# DOCUMENT RESUME

ED 043 690 UD 010 600

AUTHOR Jensen, Arthur R.; Rohwer, William D., Jr.
TITLE An Experimental Analysis of Learning Abilities in Culturally Disadvantaged Children. Final Report.
INSTITUTION California Univ., Berkeley.
SPONS AGENCY Office of Economic Opportunity, Washington, D.C.
PUB DATE Jul 70
NOTE 205p.

EDRS PRICE EDRS Price MF-$1.00 HC-$10.35
DESCRIPTORS Aptitude Tests, Associative Learning, Cognitive Ability, *Disadvantaged Youth, Educational Testing, *Ethnic Groups, Intelligence Differences, Language Ability, *Learning Characteristics, Nonverbal Ability, Nonverbal Tests, Preschool Learning, Research Methodology, Social Class, *Social Differences, Statistical Data, Test Results

ABSTRACT

This report contains a series of studies which represent ongoing research of six investigators, who seek to elucidate through empirical studies the psychological characteristics of culturally disadvantaged children. The chief aim has been to make comparative analyses of abilities and learning characteristics of children from intact subpopulation groups that differ markedly in the degree of school success typically achieved. The studies focus on: a two-level theory of mental abilities; the organization of abilities in preschool children; level I and level II performance in low and middle socioeconomic status (SES) elementary school children; relationship of the "Draw-a-Man" Test to level I and level II; comparison of "culture-loaded" and "culture-fair" tests; social class differences in free recall of categorized and uncategorized lists; mental elaboration and learning proficiency; ethnicity-SES and learning proficiency; and, elaboration training and paired associate learning efficiency in children. Appendixes contain some of the test forms used. (RJ)

# AN EXPERIMENTAL ANALYSIS OF LEARNING ABILITIES IN CULTURALLY DISADVANTAGED CHILDREN

Arthur R. Jensen

William D. Rohwer, Jr.

University of California, Berkeley

FINAL REPORT

Office of Economic Opportunity

Contract Number OEO 2404





July, 1970

Table of Contents

List of Tables

List of Tables (Cont'd)

List of Tables (cont'd)

## List of Figures

List of Figures (cont'd)

List of Figures (cont'd)

## List of Appendices

# Acknowledgements

This report results from a contract between the Office of Economic Opportunity and the University of California for a study of the abilities and learning characteristics of disadvantaged children. The work represents a continuation of the research programs of the principal investigators, Arthur R. Jensen and William D. Rohwer, Jr., to elucidate through empirical studies the psychological characteristics of the disadvantaged. While the principal methodology of these studies has involved comparisons of racial and social class groups, the chief aim has not been the study of these variables *per se*. Instead, the aim has been to make comparative analyses of abilities and learning characteristics for children from intact subpopulation groups that differ markedly in the degree of school success typically achieved.

No attempt has been made to present a theoretically unified approach to all the studies undertaken under this contract, nor would it seem desirable to do so at the present stage of our progress. The investigators have a common goal -- the scientific description and understanding of the psychology of the disadvantaged -- but each has approached the problem through his own methodology and theoretical orientation. Through such an interplay of methods, theoretical concepts, and mutual criticism, it is hoped that a unified theory of educational differences may eventually take shape. The authorship of each of the several studies has therefore been specifically indicated at this point. Each study is presented as a self-contained report, although Jensen's general introduction serves to give a unified overview of the theoretical formulations which provide the basis for his several studies and Rohwer's introductory paper performs a similar function for the studies conducted under his direction. It is intended that all the studies will duly appear in the appropriate psychological and educational journals.

The principal investigators are greatly indebted to their collaborators, whose high competence and unusual dedication was indispensable for the execution of these studies. Special credit is due Dr. Douglas Penfield, the Project Director, for so successfully coordinating all the work in the several schools in each of three districts and for supervising the initial tabulation and analysis of the data. Mrs. Mary Sue Ammon, Mrs. Janet Frederiksen, Miss Marsha Lynn, Miss Helgola Ross, and Dr. Nancy Suzuki conducted all the individual testing in the several projects. In addition, Mrs. Ammon managed the logistics of data collection in the several studies and created the instructional materials and procedures for the study of elaboration training. Dr. Steve Lynch skillfully created and produced all of the filmed PA learning materials used in several of the studies. Dr. Wade Egbert supervised the large-scale group testing, scoring, and parts of the data analysis in Jensen's final study of the relationships between Level I and Level II abilities in a total school population. Dr. Carol Treanor served as statistician and computer analyst in the studies by Jensen and Dr. Joel Levin designed and conducted the final statistical analyses for the studies by Rohwer. Mrs. Carol Dodson ably discharged the responsibilities of administrative assistant for the duration of the project. The report was typed and compiled by Miss Judy Such and Mrs. Dodson; their excellence is responsible for the polished appearance of the final product. All the figures were drafted by Mrs. Katherine Eardley and the tables were typed by Mrs. Enid Goetzl. Finally, the staffs and the children of the participating schools in Berkeley, Orinda, and Richmond have our gratitude for their unstinting cooperation; we hope their efforts will eventually receive a more sufficient recompense.

Arthur R. Jensen

William D. Rohwer, Jr.

# A Two-Level Theory of Mental Abilities

Arthur R. Jensen

Several of the studies that follow were intended to investigate a theoretical conception that had grown out of earlier work concerning the structure of mental abilities as a function of socioeconomic status (SES). The various empirical studies reported here can be more easily understood if their theoretical basis is completely spelled out in advance.

Previous research on the intelligence and learning abilities of children called culturally disadvantaged, to discover the ways in which they differ typically from middle class children in their intellectual capabilities, had led to the formulation of a theory of mental ability which can comprehend most of the phenomena revealed by these investigations (Jensen, 1961, 1963, 1968a, 1968b). The theoretical formulation has also served as a basis for predicting new phenomena concerning the relationship between intelligence, learning ability, and socioeconomic status (SES). The theory evolved gradually to accommodate our growing body of psychometric and experimental data, and it is still in a formative stage. However, it has been sufficiently formalized to yield predictions of new phenomena and to be subjected to experimental tests by other investigators. It has also been subjected recently to certain criticisms (Humphreys & Dachler, 1969; Jensen, 1969b). One aspect of the theory, at least, is still of doubtful validity, although it has not yet been put to a wholly appropriate test. Since some of the studies that led to the formulation of the theory can be better understood in light of the theory, it will be less to the reader's advantage to present this material in historical sequence than to present it in relation to the key aspects of the theory. To provide an over-view of the theory, it will be outlined first without reference to empirical evidence, which will be filled in later.

The Dimensionality of Social Class Differences.-The research literature on social class differences in intelligence makes it apparent that the evidence on social class differences in intelligence cannot be readily systematized or comprehended without positing at least two empirical dimensions along which the differences range. The work of Eells *et al*. (1951) was perhaps the most influential in arriving at this formulation, although Eells himself did not make the formulation explicit in his own work. Eells pointed out on the basis of his massive data, in which individual test items were analyzed in terms of the percentage of children in different SES groups who could answer the item correctly, that the SES differences were related to (*a*) the cultural content of the test items and to (*b*) the complexity of the items, that is, the degree of abstractness and problem solving involved in the test item. Thus, one dimension along which test items can range is that of cultural loading, by which we mean the differential probability of exposure or opportunity to become familiar with the content of the item from one social class environment to another. Test items involving knowledge of musical instruments, exotic zoo animals, and fairy tales, for example, can be said to have a high cultural loading. Whole tests can differ on this dimension of culture-fairness. Jensen has proposed that a main criterion of culture-fairness of tests be their heritability (i.e., the proportion of variance attributable to genetic factors) in the population in which they are standardized and used (Jensen, 1968c). Eells et al. (1951) also noted that the largest social class differences did not show up on the most culturally loaded items, but rather on those items that involved the highest degree of abstraction, conceptual thinking, and problem solving ability. Often these items had no cultural content to speak of, in the sense of differential exposure of item content in different social classes. Besides, if all of the SES intellectual difference were due to differences between SES groups in cultural experiences, it should be possible to devise intelligence tests that favor low SES groups over high SES groups. So far no one has succeeded in doing this. The few attempts have failed to meet a crucial criterion, namely, that the test should still correlate highly with other measures of intelligence. If lower Stanford-Binet IQs in low SES groups are due to differences in cultural experience, it should be possible to devise a test which correlates with the Stanford-Binet but which gives low SES children higher IQs than middle SES children. In other words, culture bias in tests should be completely reversible. Despite energetic efforts, no one has been able to show that this is in fact possible, which leads me to the conclusion that the culture bias factor in SES intelligence differences is indeed a real effect, but a trivial one as compared with SES differences due to abstractness and complexity of test items. Tests can be devised to minimize the culture factor, but if they are to remain intelligence tests, with the predictive validity in our society that intelligence tests are known to have, they cannot minimize the complexity factor.

Figure 1 shows this two-dimensional space, with the hypothetical

- - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - -

location of various tests in the space. The X-axis (horizontal) is the culture-loading dimension, defined by the theoretical extremes of complete heritability ($h^2 = 1$), in which there is no environmental variance in the test scores, and the other extreme of zero heritability, in which all the variance is attributable to environmental factors. The Y-axis (vertical) is the complexity dimension, going from conditioning and simple associative learning up to complex conceptual learning and abstract problem solving. Tasks can be found at every point on this continuum; tests do not fall into discrete classes. Another point that needs to be emphasized is that a particular test does not necessarily have an invariant position in this two-dimensional space. Some tasks lend themselves to being learned on an associative level or on a conceptual level, and different learners may prefer one or the other approach, so that in one population a test may stand at a different point on the complexity continuum than in another population. Paired-associate learning is not represented in Figure 1 simply because it is so ambiguous with respect to the complexity dimension. Some subjects will learn the

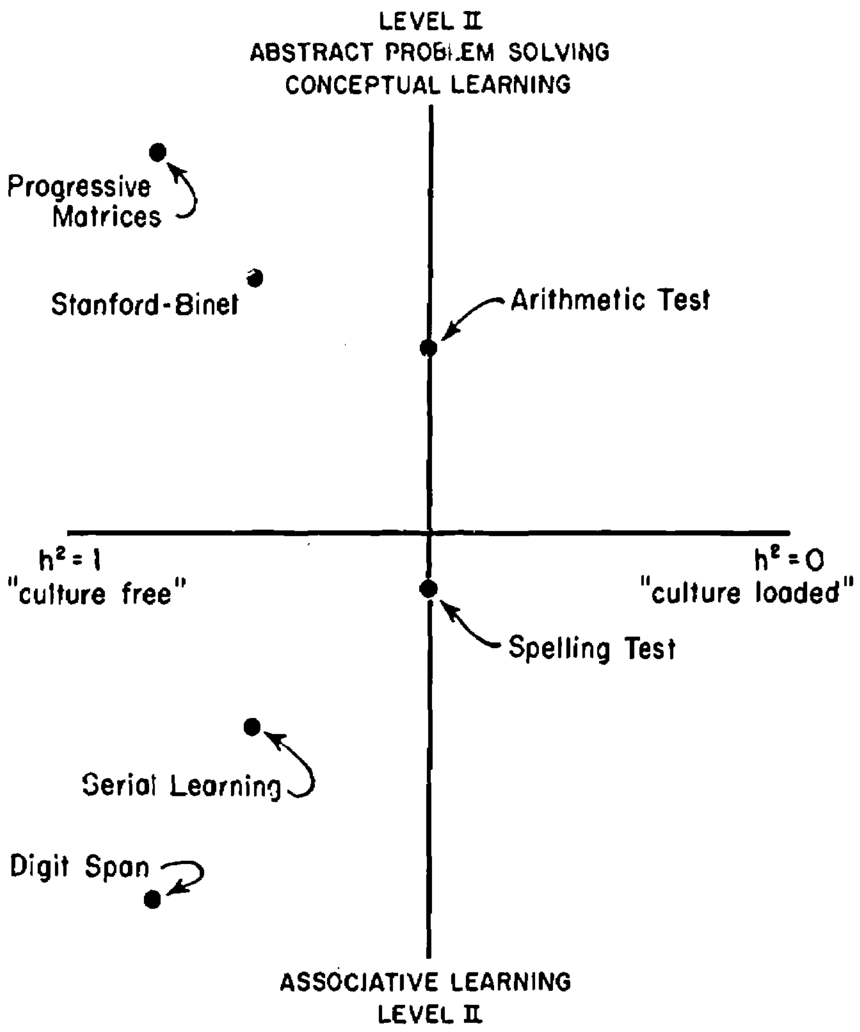



Figure 1. The two-dimensional space required for comprehending social-class differences in performances on tests of intelligence and learning ability. The locations of the various "tests" are speculative.

pairs by rote, others by means of conceptual mnemonic processes, depending upon the age and pattern of abilities of the subjects. Other tasks, like digit span and serial rote learning, are much less flexible in this respect, and nearly always stand low on this continuum. At the other extreme, complex tasks like Raven's Progressive Matrices cannot be solved by simple associative processes and are therefore relatively fixed near the upper end of the continuum.

Although tests range continuously along this dimension, the dimension itself is viewed theoretically as being the result of two different types of mental ability which can be distributed independently in a given population. In other words, the diagram in Figure 1 is intended to describe phenotypic test performance and not the underlying genotypic abilities which find expression through these various tests.

Genotypic Abilities: Level I and Level II.-The Y-axis in Figure 1 represents the relative admixture in various tests of two fundamental genotypes of ability, which I call Level I (associative learning ability) and Level II (conceptual learning and problem solving). By "genotype" I mean simply the physiological substrate of the ability, regardless of whether it is genetically or experientially conditioned. The vertical axis in Figure 1 can be resolved into two dimensions, Level I and Level II. Points along the vertical axis in Figure 1 can be thought of as lying on various vectors in the two-dimensional space created by the Level I and Level II dimensions.

Level I ability is essentially the capacity to receive or register stimuli, to store it, and to later recognize or recall the material with a high degree of fidelity. Jensen (1968a) originally called Level I "basic learning ability." It is characterized especially by the lack of any need for elaboration, transformation, or manipulation of the input in order to arrive at the output. The input need not be referred or related to other past learning in order to issue in effective output. A tape recorder exemplifies Level I ability in its most extreme and pure form. In human performance forward digit span is one of the clearest examples of Level I ability. Reception and reproduction of the input with high fidelity is all that is required. Reverse digit span would represent a less pure form of Level I ability, since some transformation of the input is required prior to output. Serial rote learning and paired-associate rote learning, especially when the stimulus and response items are relatively meaningless and thereby do not lend themselves very much to verbal mediation or transfer from prior verbal learning, are largely dependent upon Level I ability. Level I is seen as the source of most individual differences variance in performance on rote learning tasks, digit span, and other types of learning and recall which do not depend upon much transformation of the input.

Level II ability, on the other hand, is characterized by transformation and manipulation of the stimulus prior to making the response. It is the set of mechanisms which make generalization beyond primary stimulus generalization possible. Semantic generalization and concept formation depend upon Level II ability; encoding and decoding of stimuli in terms of past experience, relating new learning to old learning, transfer in terms of concepts and principles, are all examples of Level II. Spearman's characterization of g as "eduction of relations and correlates" corresponds to Level II. Most standard intelligence tests, and especially so-called culture-fair tests such as Raven's Progressive Matrices and Cattell's Culture Fair Tests of g, depend heavily upon Level II ability. Since Level I ability is needed for high fidelity reproduction and is thus exemplified by a tape recorder, Level II ability is needed for transformation and elaboration of stimulus-response elements and what Spearman would call the fundaments of learning and is thus exemplified by the intellectual performance of a Newton and a Beethoven, who performed elaborate transformations on clearly circumscribed symbol systems -- mathematics and music.

Few if any tests tap either Level I or Level II in a pure form, but some tests depend much more upon one than upon the other. Persons tend to use the abilities they've got, and so we find some subjects approaching what for most subjects is a Level I task as if it were a Level II task. At times this can result in poorer performance on a task. We have had bright college students, for example, approach a task which could be learned only by rote (since it involved only a random pairing and reinforcement of stimulus-response contingencies) as if it were a logical problem-solving task; their attempts to "break the code" of what was only a random sequence of stimuli actually delayed their mastery of the task, a mastery which average young school children attained considerably faster, since only their Level I ability was brought to bear upon it.

Level I and Level II abilities are seen as largely genetically conditioned. The heritability of high Level II tests, such as the Progressive Matrices, is already clearly established, and there is no reason to suppose that Level I tests would not have equally high heritability (Jensen, 1967b, 1968a, 1968c, 1969a, 1969b). But the exact heritability of Level I and II is not so important, in terms of our theory, as the postulation that the mechanisms of Level I and II are genotypically independent. They may be correlated in any given population, but since, according to the theory, they are due to genetic factors which can be assorted independently, they need not be correlated. Correlation can come about in two ways: (a) through genetic assortment of the two types of ability and (b) from a hierarchical functional dependence of Level II upon Level I. But discussion of these points should be postponed until a few more basic issues have been explicated.

Hierarchical Dependence of Level II Upon Level I.-Level II processes are viewed as functionally dependent upon Level I processes. This hypothesis was formulated as a part of the theory to account for some of our early observations that some children with quite low IQs (i.e., 50 to 75) had quite average or even superior scores on Level I-type tests (simple S-R trial and error learning, serial and paired-associate rote learning, and digit span), while the reverse relationship did not appear to exist: children who were very poor on the Level I tests almost never had high IQs. It also seems to make sense psychologically to suppose that basic learning and short-term memory processes are involved in performance on a complex Level II task, such as the Progressive Matrices, although the complex inductive reasoning strategies called for by the matrices would not be called upon for success in Level I tests such as digit span and serial rote learning. Therefore it was hypothesized that Level II performance depends upon Level I but not vice versa. In other words, Level I ability is seen as necessary-but-not-sufficient for the manifestation of Level II ability. A person who was very deficient in Level I would never manifest high Level II ability -- even if his genotype for Level II were in the superior range. On the other hand, an individual's Level I ability could be manifested on many tasks irrespective of his endowment of Level II ability. This kind of functional dependence of Level II upon Level I implies a "twisted pear" type of correlation between tests that represent each of these levels. Of course, if tests of Level I and Level II were constructed so as to yield a normal distribution of scores in the total population, a bivariate normal scatter diagram would be forced on the data and the "twisted pear" would be constrained from appearing. Since there is already good evidence that Level II, as measured by standard intelligence tests, is approximately normally distributed in the population, we would hypothesize that Level I functions have a positively skewed distribution. So far, however, we have no compelling evidence on the shape of the distribution of scores on Level I tests, such as digit span, in the general population. Investigation of the hypothesized functional dependence of Level II upon Level I can probably best be determined from the study of neurological evidence. No thorough study of this nature has yet been attempted. Some evidence indicates that brain damage and aging which affect Level I processes (short-term memory, etc.) also depress performance on Level II tests such as the Progressive Matrices (Horn, 1970), although the reverse does not seem to hold -- Korsakow patients, for example, show defects in conceptual reasoning and problem solving but have digit spans within the normal range (Talland, 1965). On the other hand, exceptionally high Level I abilities, such as Luria (1968) described in a man who could memorize more than 100 items in a serial or paired-associate list in a single trial, are not necessarily accompanied by a high level of ability in abstract, conceptual reasoning. Luria's subject, in fact,had quite mediocre conceptual abilities. These findings suggest the necessary-but-not-sufficient relationship of Level I to Level II.

Distributions of Level I and Level II as a Function of Socioeconomic Status (SES).-The theory postulates that Level I ability is about equally distributed in all SES groups. In short, there is little, if any, correlation between Level I ability and SES. On this point the theory will probably have to be modified slightly, so that there will be a low positive correlation between Level I and SES. To keep the theoretical formulation as simple as possible for the purpose of explication, however, we will posit no SES difference in Level I.

Level II ability is distributed quite differently as a function of SES, there being a positive correlation between Level II and SES. Figure 2 shows the hypothetical distributions of Levels I and II in lower-class and middle-class populations.

- - - - - - - - - - - - - - -
Insert Figure 2 about here
- - - - - - - - - - - - - - -

Why are these abilities said to have different distributions in lower and middle-class segments of the population? It can be argued that the educational and occupational requirements of our society tend to sort people out much more by their Level II ability than by their Level I ability, and it is occupational status that chiefly determines an individual's SES. Assuming largely genetic determination of individual differences in both Levels I and II, the "gene flow" would diffuse in both directions with respect to SES. If Level II is dependent upon Level I, then high SES children who are low on either Level I or II will tend as adults to gravitate to a lower SES level. If their deficiency is at Level I only, they will carry good genes for Level II with them in many cases; if their deficiency is only at Level II, however, they will carry good genes for Level I with them as they gravitate to a lower SES. Moving from lower to higher SES, on the other hand, carries with it good genes for both Level I and Level II. This set of conditions is consistent with two well-established sets of observations. Kushlick (1966, p. 130), in reviewing the research on SES and mental subnormality, notes that cultural-familial retardation (IQs between 50 and 75) is predominantly concentrated in the lower social classes. On the basis of a number of surveys made largely in England, Kushlick concludes that mild subnormality is the absence of abnormal neurological signs and is virtually confined to the lower social classes. He goes on to say that almost no children of higher social class parents have IQ scores less than 80, unless they have a pathological condition. In short, genes for low intelligence (meaning low Level I and/or low Level II, according to our theory) are largely eliminated from the upper SES segment of the population. (Severe mental deficiency, due to brain damage and mutant gene and chromosomal defects, however, have about equal occurrence in all social strata.) The second important observation that is consistent with our formulation is the fact that it is not nearly as difficult to find gifted (IQs above 130) children in the lower classes as it is to find retarded children in the upper classes. The Scottish National Survey established on a large scale that high intellectual ability is more widely distributed over different social environments than is low mental ability (Maxwell, 1953). This is what we should expect if many genes for high Level II ability gravitated from upper to lower classes as a result of having been combined with poor Level I ability. In reassortment the good Level II genes can combine with good Level I genes to produce a high level of general ability, which then will tend to be upwardly mobile in the SES hierarchy.

<u>Level I-Level II Correlation in Low and Middle SES</u>.-From the foregoing considerations we can propose a crude model that "predicts" the form of the correlation scatter diagram between Level I and Level II tests. We begin with the hypothetical distribution of genotypes for Level I and Level II in lower and middle SES. Assume that we divide each of these distributions at the common median for the total population, as follows:

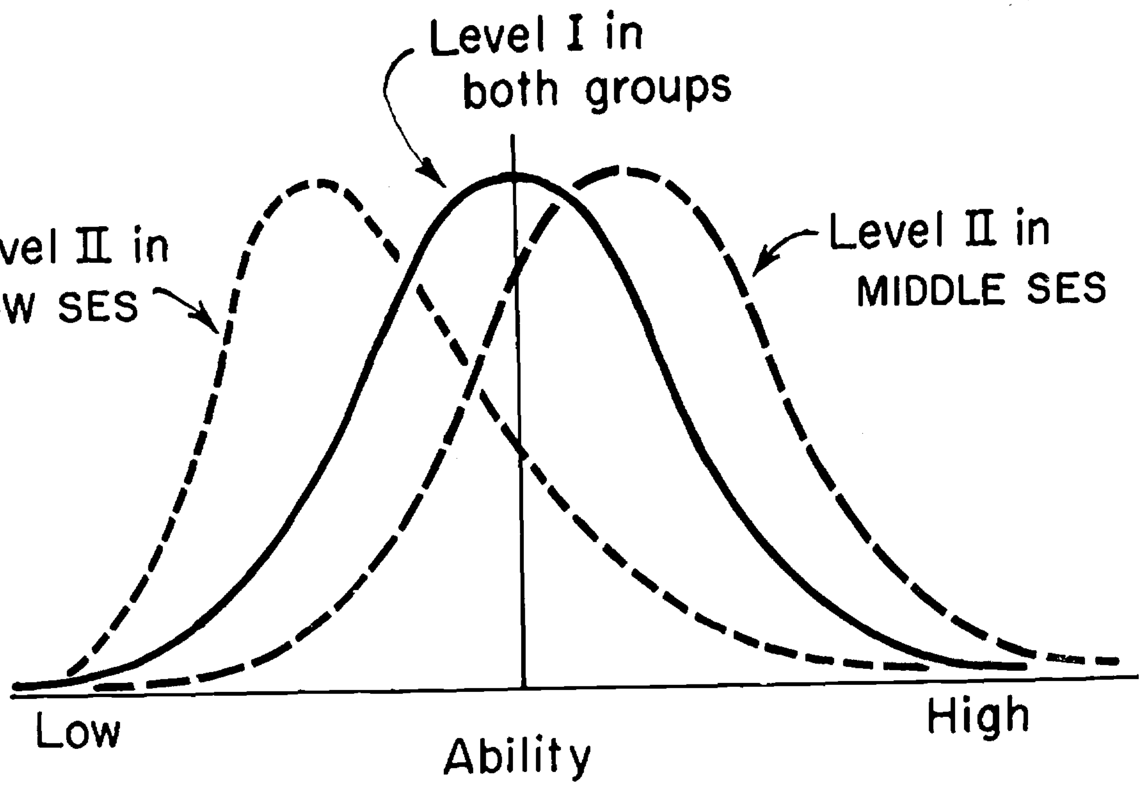
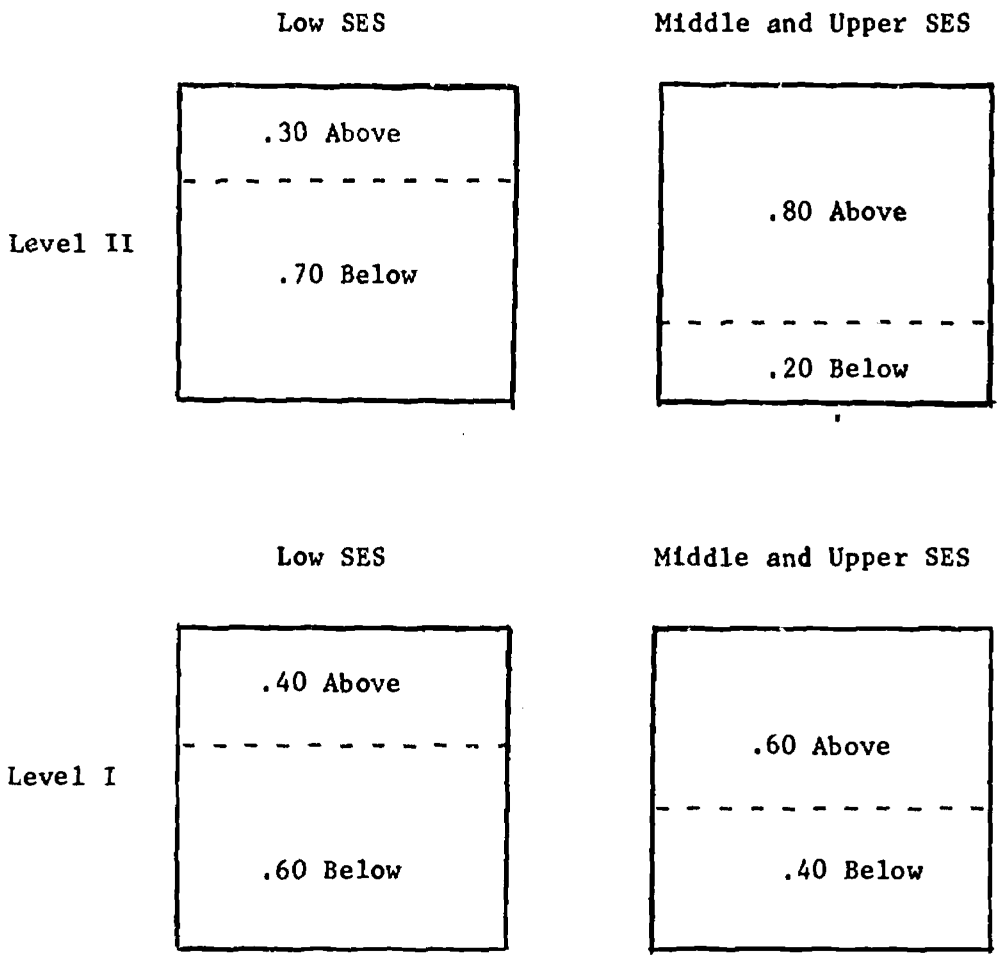



Figure 2. Hypothetical distributions of Level I (solid line) and Level II (dashed line) abilities in middle-class and lower-class populations.




Phenotypes on Level I and Level II tests are produced by the joint action of individual's genotypic standing on each Level. To keep the model simple, we will say that within each social class Level I and Level II genotypes are uncorrelated, so that the proportion of phenotypes that fall above and below the population median can be obtained simply from the product of the independent probabilities of the genotypes. This is shown in the contingency tables below. The entries within the cells represent proportions of <u>genotypic</u> combinations of Level I and Level II; the marginal totals represent the proportions of phenotypes on Level I and Level II tests.

Low SES

| | | Level II Below | Level II Above | |
|---|---|---|---|---|
| Level I | Above | .28 | .12 | .40 |
| | Below | .42<br>(.18)← | (0)<br>(.18) | .60 |
| | | .88 | .12 | |

Middle and Upper SES

| | | Level II Below | Level II Above | |
|---|---|---|---|---|
| Level I | Above | .12 | .48 | .60 |
| | Below | .08<br>(.32)← | (0)<br>(.32) | .40 |
| | | .52 | .48 | 1.00 |

Genotypes in quadrant 4 are shown in parentheses, since their phenotypic performance will be much like that of subjects in quadrant 3, because of the assumed functional dependence of Level II performance on Level I ability. Thus the proportion in quadrant 4 is shown by the arrow as being moved into quadrant 3 in order to arrive at the total proportions of phenotypes. Leaving zero frequency in quadrant 4 is, of course, an overly idealized situation. Because the degree of dependence of Level II performance on Level I is far from complete, there will actually be some subjects remaining in quadrant 4, and we can hypothesize that with increasing age of subjects, from early to late childhood, we should see "late bloomers" moving from quadrant 3 to quadrant 4, with the growth of Level II functions. These intellectual late bloomers will be children with relatively low Level I ability and relatively high Level II. Thus the incidence of low phenotypic ability would be expected to decrease with increasing age of the subject population, and much more so in the middle than in the lower SES group.

According to this formulation, the correlation scatter diagrams between Level I and Level II tests would appear somewhat as is shown in exaggerated form in Figure 3. The "twisted pear" is most evident in the Low SES group, with many subjects in quadrant 1, i.e., above

- - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - -

average in Level I and below average in Level II. The model clearly predicts a much lower correlation between Level I and Level II tests in the Low SES segment of the population than in the middle SES segment. It is an empirical fact that these correlations differ in the way depicted by the model, which was devised to account for the difference in correlations between Level I and Level II in lower and middle-class groups. The difference in correlations cannot be accounted

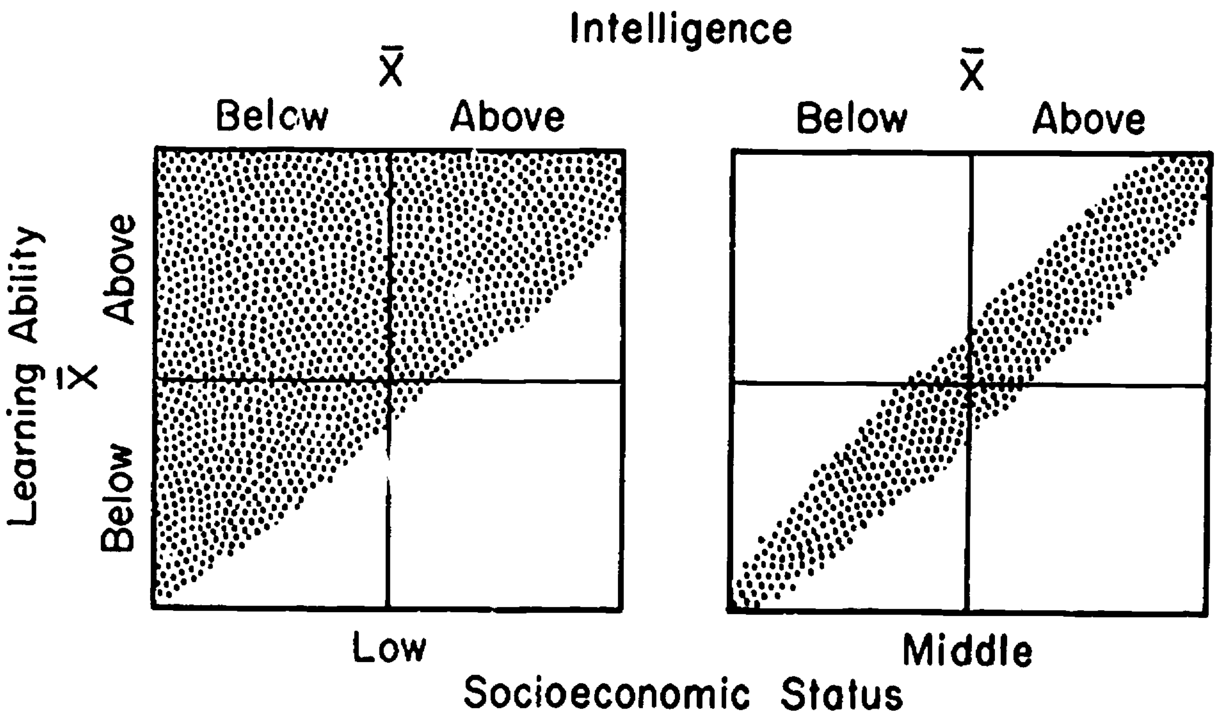



Figure 3. Schematic illustration of the hypothetical forms of the correlation scatter-diagram for the relationship between Level I (e.g., digit span) and Level II (e.g., IQ) abilities in low and middle SES groups.

for by restriction of range in the low SES group or by differences in test reliability. A theory of intelligence must be able to account for the well-established difference in correlations. The present model does so and is also consistent with much other evidence. At present, however, the model can only be regarded at best as a rather crude first approximation to the model that will hopefully evolve as a result of empirical investigations directed at obtaining the kinds of information needed for refining the model and rigorously testing its basic assumptions.

Growth Curves of Level I and Level II Abilities.-It is hypothesized that Level I and Level II have quite different growth curves, as shown in Figure 4.

- - - - - - - - - - - - - - -

Insert Figure 4 about here

- - - - - - - - - - - - - - -

No scale is indicated on the Y-axis and therefore the exact shape of the growth curves should not be taken too literally. They are merely intended to convey the hypothesis that Level I rises rapidly with age, approaches its asymptotic level relatively early, and shows little SES difference, as contrasted with Level II, which does not begin to show a rapid rise until 4 or 5 years of age, beyond which the SES groups increasingly diverge and approach quite different asymptotes. The forms of the Level I and Level II curves express some of the developmental characteristics that White (1965) called associative ability (Level I) and cognitive ability (Level II). The hypothesis shown in Figure 4 has clear predictive implications for the magnitude of SES differences as a function of age and of type of test.

## Previous Empirical Evidence

Most of the empirical data relevant to the theory has already been presented elsewhere and is only summarized here. The earlier studies produced the phenomena which the theory has been devised to explain and were not designed as tests of the theory. Later studies, however, have grown out of deductions from the theory and were designed to test specific hypotheses.

Independence of Level I and Level II.-If Level I phenotypes are defined by scores on digit span and laboratory measures of rote learning, and Level II is defined by scores on standard intelligence tests, particularly those with the highest *g* loading, such as the Progressive Matrices, and by laboratory tasks involving conceptual learning and abstract problem solving, there is ample evidence that these two classes of tasks, Level I and Level II, are factorially distinct abilities. As indicated in our theoretical formulation, they are phenotypically more distinct in lower than in upper SES populations, due to the positive assortment of genotypes and to the hierarchical dependence of Level II upon Level I. In high SES groups there will be a substantial *g* loading on both Level I and Level II

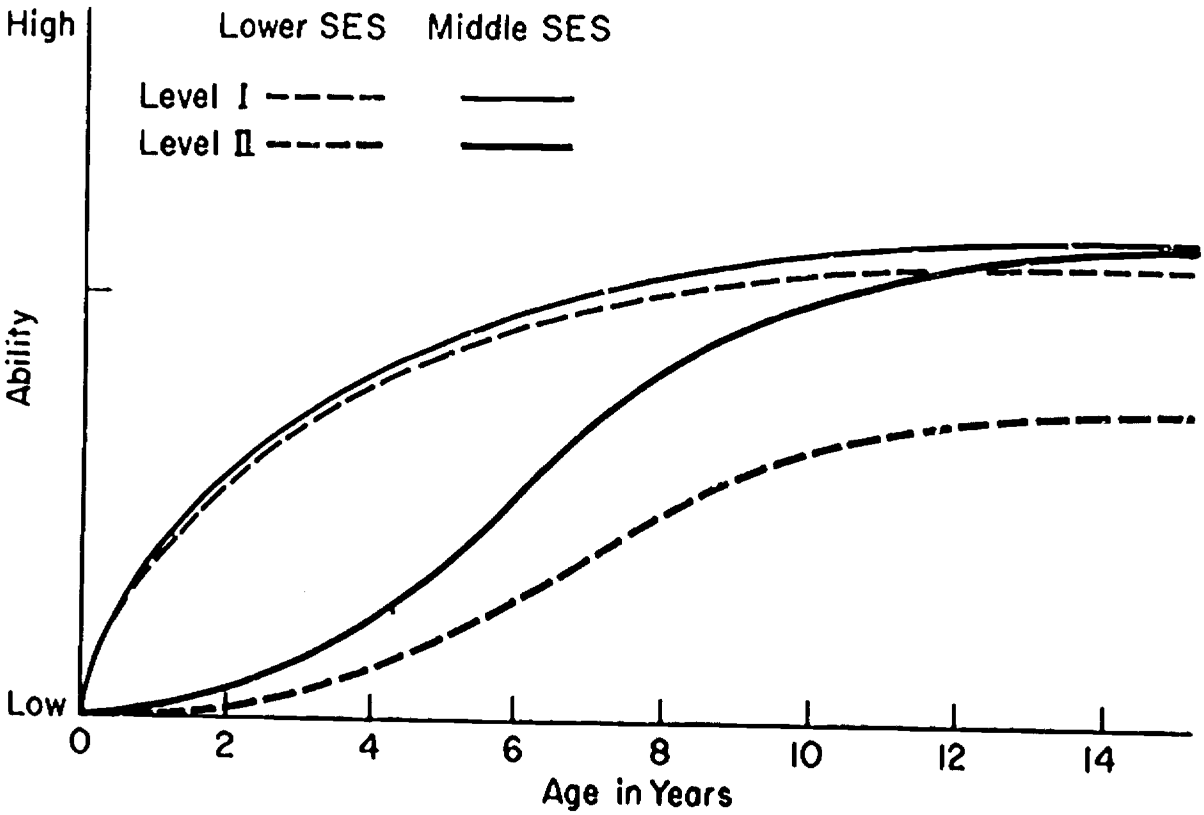



Figure 4. Hypothetical age growth curves for Level I and Level II abilities in low and middle SES populations.

tests. The fact that very low correlations are found between the two types of tests in some population groups, however, argues for their factorial independence. Zeaman and House (1967) have reviewed the research relating IQ to learning abilities, which shows, in general, that as the learning task becomes more rote, it correlates less with IQ. As learning tasks increase in discriminative and conceptual complexity (not necessarily in difficulty) they are more highly correlated with IQ. Even reverse digit span, since it involves a transformation of the stimulus input, is more highly correlated with *g* than is forward digit span (Horn, 1970).

Triple Interaction of IQ, Learning Ability, and SES.-The early studies focused on the interaction of IQ, learning ability, and SES. The basic design of these studies was a 2 X 2 analysis of variance, with Low vs. High IQ on one dimension and Low vs. High (or Middle) SES on the other. In three of the studies (Jensen, 1961, 1963; Rapier, 1966) the low IQ subjects were in special classes for the educable mentally retarded. This particular experimental design has been criticized by Humphreys and Dachler (1969a, 1969b) on the grounds that it is "pseudo-orthogonal," i.e., it treats IQ and SES as if they were uncorrelated in the population by having equal Ns in the four cells of the 2 X 2 analysis of variance. Unless the results are manipulated by weighting the cell means proportionally to the frequencies of the groups in the population, the results of the analysis can be said to be biased, that is, they cannot be generalized to the total population. Jensen (1969b) argued in turn that the pseudo-orthogonal design served legitimately to disclose the existence of an interaction between IQ, learning ability, and SES and could now be followed up by correlational studies in representative population samples to establish the magnitudes of these intercorrelations.

The essential features of the data of these early studies are shown in Figure 5. The low SES groups in the studies summarized in

- - - - - - - - - - - - - - -
Insert Figure 5 about here
- - - - - - - - - - - - - - -

Figure 5 have been either white children (Rapier, 1966), Mexican-American children (Jensen, 1961) or Negro children (Jensen & Rohwer, 1968). The findings are essentially the same regardless of race, though it should be noted that in selecting groups of children who are high or low on SES and above or below average in IQ, our samples represent different proportions of each racial population. The groups labeled High-SES in these studies were in all cases white middle or upper-middle-class children.

Figure 5 shows a marked interaction between SES, IQ, and learning ability of the type measured by tasks of free recall, serial learning, paired-associates learning, and memory for digit series. Low SES children in the IQ range from 60 to 80 perform significantly better in these learning tasks than do middle-class children in the same range of IQ. Low SES children who are average or above average in IQ, on

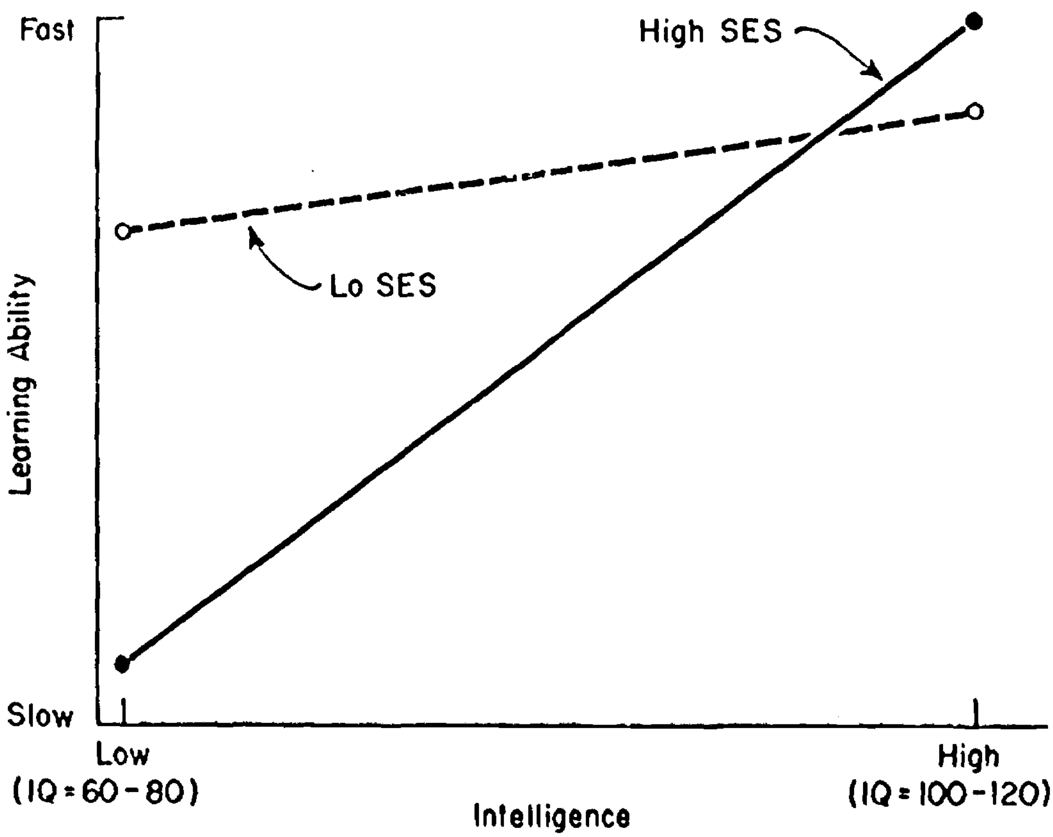



Figure 5. Summary graph of a number of studies showing relationship between learning ability (free recall, serial and paired-associate learning) and IQ as a function of socioeconomic status.

the other hand, did not show learning performance that is significantly different from that of middle-class children of similar IQ in these early studies.

The theory has been made to predict this interaction, so it should not be surprising that these data fit the theory. Since the formulation of the theory, however, this interaction has been predicted in new data. Durning (1968) designed a study specifically to test several hypotheses derived from the theory. She obtained data on 5,539 Navy recruits (". . . approximately the total input for a period of six weeks to the Naval Training Center, San Diego"); 95 percent of them were between 18 and 23 years of age, with an average of 11.9 years of school. They were given a battery of standard selection tests, including the Armed Forces Qualification Test (AFQT), and a special auditory digit memory test, with a reliability of .89. Durning predicted, in accord with the present theory, that Negro recruits who scored low on the selection tests would obtain higher digit memory scores than non-Negro recruits with low scores on the selection tests. She compared Negroes and non-Negroes in Category IV (AFQT scores between the 10th and 30th percentiles), and concluded: "Negro CAT-IVs as a group scored significantly higher on the Memory for Numbers Test than non-Negro CAT-IVs, though the Negroes were lower on most of the standard selection tests" (Durning, 1968, p. 21). CAT-IV recruits, especially Negroes, come largely from low SES and culturally disadvantaged backgrounds.

SES Differences on Level I and Level II.--In every study we have performed it has been found that low-SES and middle-SES groups differ much less on Level I tests than on Level II. Jensen (1963) found some low SES children with Stanford-Binet IQs in the range from 50 to 75, who on a Level I test (trial-and-error selective learning) exceeded the mean performance of children of the same age classed as "gifted" (IQs above 135). None of the gifted, however, scored below average children (IQs 90-110).

Groups of normal children selected at random from regular classes in grades K (kindergarten and Head Start classes), 1, 3, and 6 were given a paired-associates test devised by Rohwer, using picture pairs presented by means of a motion picture projector. The children were sampled from populations of low and middle SES. These groups differ by 15 to 20 points in IQ. Included in the study was a group of 48 institutionalized familially retarded young adults; they were tested to obtain evidence that the paired-associate learning test indeed taps an important aspect of mental ability, and it was hypothesized that institutionalized retardates would be deficient in Level I as well as Level II ability (Jensen & Rohwer, 1968). Figure 6 shows the results,

- - - - - - - - - - - - - - -
Insert Figure 6 about here
- - - - - - - - - - - - - - -

which indicate that the learning test shows a significant age trend but no significant SES difference. Furthermore, the adult retardate group is lower than any other group in the study and significantly

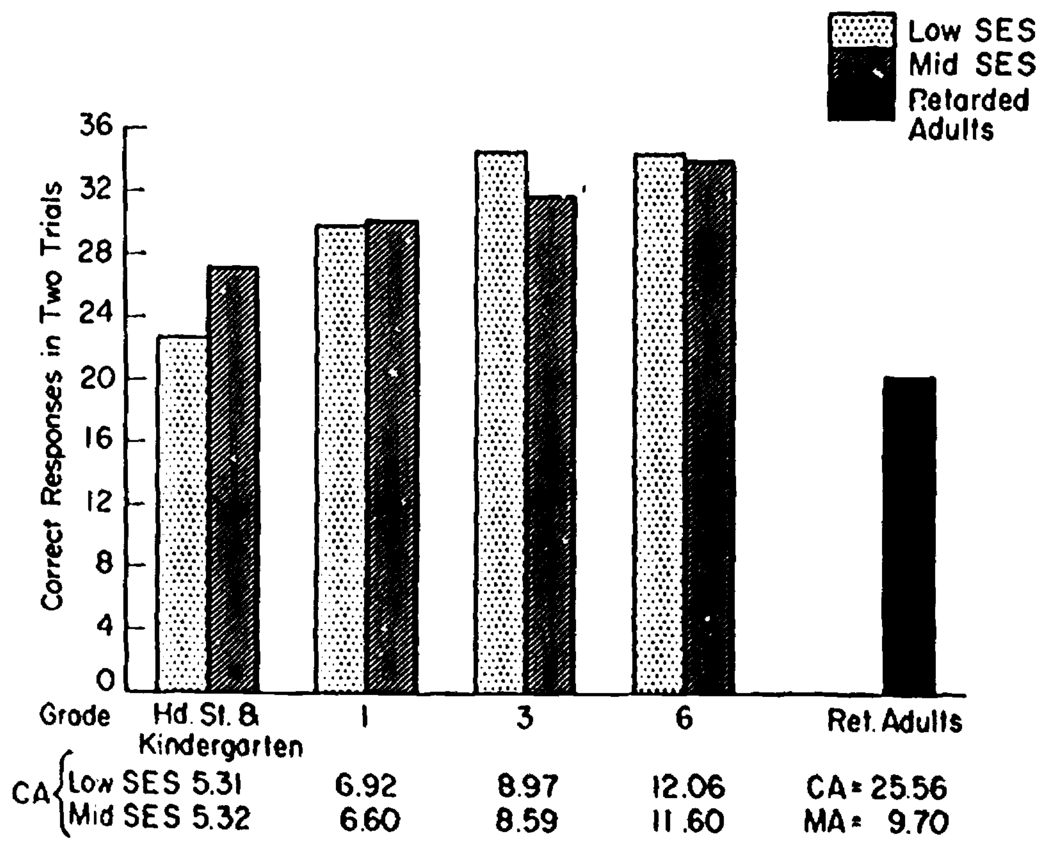



Figure 6. Comparisons of low and middle SES groups of children at various grades in school with institutionalized retarded adults on paired-associate learning consisting of 24 picture pairs presented for two trials at a rate of 4 seconds per pair. $\underline{N}$ = 48 in each of the nine groups.

lower than all the other groups combined. Comparison of the learning performance of the adult retardates and the middle-SES third-graders is especially interesting, since the two groups have approximately the same mental age (9.7 versus 9.6). It is clear that in these samples the paired-associate learning is more highly related to IQ than to mental age.

In another study, Rohwer (1969) administered the Peabody Picture Vocabulary Test (PPVT), Raven's Colored Progressive Matrices, and a paired-associates learning test to a total of 288 children drawn in equal numbers (N = 48 per group) from Kindergarten, 1st and 3rd grades in two kinds of schools -- ones serving a low-SES Negro area and ones serving an upper-middle-class white residential area. The results are shown in Figure 7; to facilitate comparisons the raw test scores

Insert Figure 7 about here

were converted to T scores with a mean of 100 and a standard deviation of 15. Note that, in accord with our theory, the Negro-white or low-SES vs. high-SES difference is much smaller for the Level I (paired-associate) test than for either the PPVT or the Raven, which are both Level II tests. The Raven Matrices is presumably less culturally loaded than the PPVT. Also note that in accord with our hypothesis that SES groups diverge on Level II with increasing age (shown in Figure 4), the Negro and white groups show an increasing difference with advancing school grade on the two Level II tests, especially on the Raven. Just the reverse appears to be true for the paired-associates test.

Guinagh (1969) tested low-SES Negro (N = 105), low-SES white (N = 84), and middle-SES white (N = 79) third-graders on Raven's Colored Progressive Matrices and a digit span test. The low and middle SES groups, though differing very significantly on the Progressive Matrices, did not differ significantly on digit span.

Scholastic tests which involve more rote learning than reasoning also correlate less highly with indices of pupils' SES. For example, Project TALENT data on a 10 percent sample of male 12th graders (N = 2.946) show multiple correlations between a number of SES indices and Level II-type scholastic tests of .53 (Information), .44 (English), .46 (Mathematics), .41 (Mechanical Reasoning) as compared with only .24 for Memory for Words ("the ability to memorize foreign words corresponding to common English words") (Flanagan & Cooley, 1966, p. E-8).

Correlations Between Level I and Level II in Low and Middle SES Groups.-We have found substantial correlations between Level I tests (serial and paired-associate learning, free recall, and memory span) and IQ or MA (mental age) in middle-class children, but very low correlations in low-SES groups, as would be predicted from the forms of the scatter diagrams hypothesized in Figure 3. For example, in a study of white children, ages 8 to 13, Rapier (1966) found that the

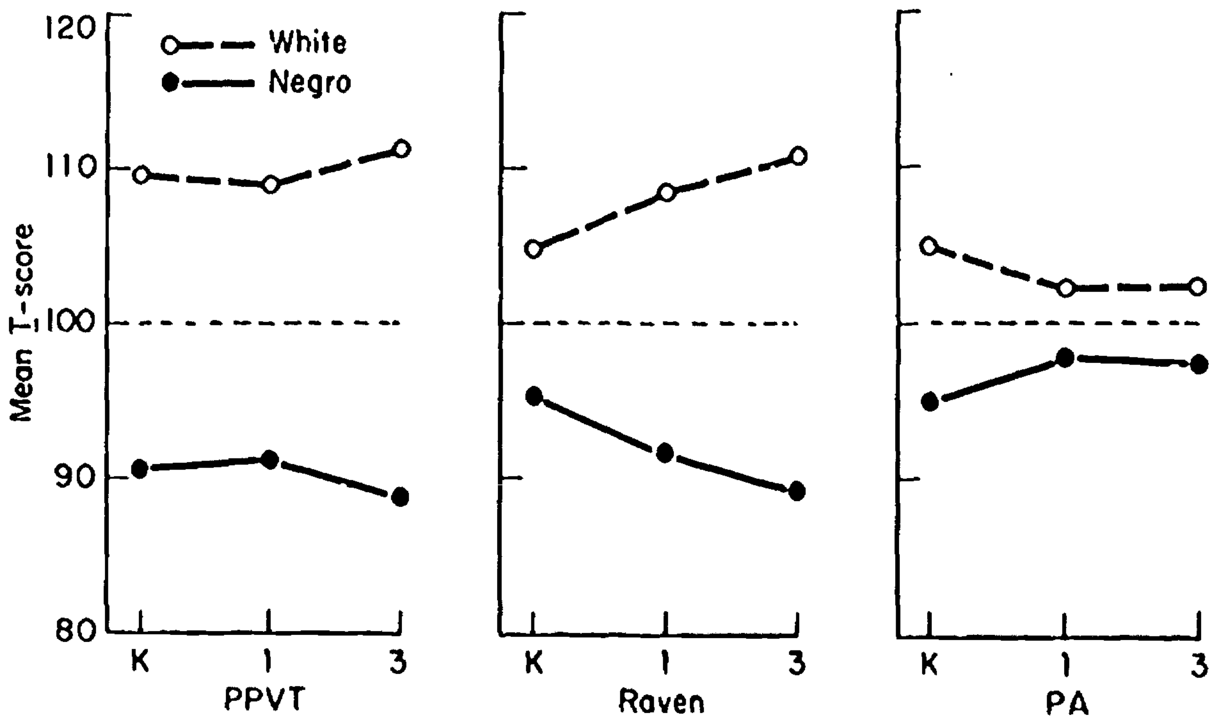



Figure 7. Performance on the Peabody Picture Vocabulary Test (PPVT), Raven's Colored Progressive Matrices, and a picture paired-associates learning test, in T scores, with mean = 100, SD = 15. (From Rohwer, 1969)

average correlation (Pearson $r$) between IQ (PPVT) and serial and paired-associate learning tasks was .44 for the middle-SES ($N = 40$) and .14 for the low-SES group ($N = 40$). Corrected for attenuation, these correlations are .60 and .18, respectively.

Guinagh (1969) obtained the following correlations (corrected for attenuation) between digit span and Progressive Matrices among third-graders: .29 for low-SES Negro ($N = 105$), .13 for low-SES white ($N = 84$), and .43 for middle-SES white ($N = 79$). An interesting finding of Guinagh's study was that low-IQ/low-SES Negro children with low digit span scores showed no significant improvement on Progressive Matrices after a specific instructional program on this type of problem solving, while low-IQ/low-SES Negro children with high digit span scores showed a significant gain on matrices performance after instruction, with the gains measured against no-instruction matched control groups.

Horning (1968), analyzing data on 5,539 Naval recruits, determined the correlation between the Armed Forces Qualification Test (AFQT) (a test of general intelligence and scholastic skills) and a digit memory test. The correlation (corrected for restriction of range) for Category IV recruits (AFQT scores between the 10th and 30th percentiles) was .21; for non-CAT-IVs it was .40, a difference significant beyond the .01 level.

Practical Validity of Level I Tests

If we have discovered a class of mental abilities (Level I) on which social class differences are much less than those found on IQ tests, it raises the question of whether it is possible to devise instruction in basic scholastic skills in such a way as to be less dependent upon Level II abilities and more fully utilize the Level I abilities which children called disadvantaged possess to a relatively greater degree. Can instruction geared to Level I ability improve the scholastic performance of the majority of low-SES children who now perform relatively poorly in school? School success is highly predictable from standard IQ tests. But is this true mainly because instruction is aimed so strongly at Level II ability? Is it necessary that a child who is low on Level II ability, but high on Level 1, fail to acquire the basic skills in school? Children who are above the general average on Level I abilities, but below the average on Level II performance, usually appear bright and capable of normal learning and achievement in many situations, although they invariably have inordinate difficulties in school work under the traditional methods of classroom instruction. Many such children who are classed as mentally retarded in school later become socially and economically adequate persons when they leave the academic situation. On the other hand, children who are much below average on Level I, and consequently on Level II as well, appear to be much more handicapped in the world of work. One shortcoming of traditional IQ tests is that they make both types of children look much alike. We therefore need tests that will reliably assess both Level I and Level II separately. Even more important is the need for research on more effective utilization of Level I ability in scholastic instruction. It seems sensible that instruction should be based upon a pupil's strengths rather than upon his weaknesses, and we have found that many children lacking strength in Level II possess strength in Level I. At present we do not know how to teach to Level I ability. Although Level I is manifested in rote learning, it is not advocated that simple notions of rote learning be the model for instruction. Instructional techniques that can utilize the abilities that are manifested in rote learning are needed, but this does not necessarily imply that the instruction consist of rote learning *per se*. We also need to find out to what extent Level II abilities can be acquired or simulated by appropriate instruction to children who possess good Level I ability but are relatively low on Level II as assessed by IQ tests. Guinagh's (1969) finding that low-SES Negro children with low IQs, but who had above average digit span (Level I), were able to improve in matrices performance after appropriate instruction seems extremely important. It should be followed up intensively.

The only study of the practical predictive validity of a Level I test (digit memory) is Durning's (1968) investigation of naval recruits. Durning correlated a battery of standard selection tests, as well as a digit memory test, with a measure of recruits' response to the first eight weeks of basic training. This measure was obtained by means of an objective paper-and-pencil test called the Recruit Final Achievement Test (RFAT). RFAT items cover basic seamanship, military courtesy and conduct, first-aid and safety, and other topics included in the eight weeks of recruit training. Durning states: "The fact that the RFAT is essentially an academic criterion is one of the major limitations of the present study, for the digit span test was chosen as a promising predictor of more practical, less scholastic criteria." Omnibus aptitude tests, such as the General Classification Test and the AFQT, correlated with the RFAT criterion in the range of .55 to .71. The verbal tests had the higher validities. Digit span correlated significantly with RFAT ($r = .30$, $p < .001$). This is not an impressive correlation, but it should be remembered that the RFAT was academically oriented. Durning concluded that ". . . though the Memory for Numbers Test was not an efficient predictor of RFAT, it nonetheless may have promise as a predictor of more practical, less academic measures of success in the Navy." Navy psychologists have since been analyzing these data further and are finding that for certain job categories within the Navy, the Memory for Numbers Test is a better predictor of success than the more academically oriented tests in the selection battery.

The theory presented here may provide a broad base for the discovery of aptitude X training interactions that will possibly prove fruitful for improving the education of many children who under present methods of instruction seem to derive little educational benefit from schooling. Present day schooling is highly geared to conceptual modes of learning, and this is suitable for children of average and superior Level II ability. But many children whose weakness is in conceptual ability are frustrated by schooling and therefore learn far less than would seem to be warranted by their good Level I learning ability. A certainly important avenue of exploration is the extent to which school subjects can be taught by techniques which depend mostly upon Level I ability and very little upon Level II. After all, much of the work of the world depends largely on Level I ability, and it seems reasonable to believe that many persons can acquire basic scholastic and occupational skills and become employable and productive members of society by making the most of their Level I ability. However, it would seem unwise at this point to recommend educational practices based on a theory that is not yet proven and has hardly begun to be explored for its specific educational implications.

The following studies were intended as further examination of the theoretical formulations set out above.

# The Organization of Abilities in Preschool Children

Arthur R. Jensen

This study is aimed at determining the relationship between rote learning and memory abilities, on the one hand, and psychometrically measured intelligence, on the other, in lower class and middle class preschool children. The theoretical formulation of the organization of mental abilities as a function of social class, explicated in the previous section, leads to the following hypotheses regarding the present study:

1. Mean differences between middle and low SES groups are greater for intelligence measures than for learning and memory measures.
2. There is a larger general factor among learning and intelligence test measures in middle than in low SES children.
   a. Zero-order correlations between learning tasks and IQ (or MA) are higher in the middle than in the low SES groups.

## Method

### Subjects

The sample consisted of 200 preschool children varying in age from 36 to 65 months, all of whom were enrolled in nursery schools of the parent cooperative variety.

Low SES Group. Half the children (*N* = 100) were from homes in which the modal occupation of the head of the household was "unemployed." All were Negro. All the families in this group were receiving public welfare financial assistance at the time the study was conducted. The group is, therefore, quite typical of children called culturally disadvantaged and who are eligible for Headstart and other compensatory programs. Despite the fact that the low SES group is Negro, the hypotheses tested in this study pertain to social status differences rather than race differences *per se*. Obviously race and SES are completely confounded in this study. Its aim, however, is to discover the characteristics of mental abilities among children who are typical of those for whom Headstart is primarily intended, and our "control" comparison groups are children selected from segments of the population that typically do well in school. Previous studies indicate that the theory put forth here could just as well be tested by comparing low and middle SES groups in racially homogeneous samples. Social class is not postulated to be a *causal* variable in the determination of mental abilities, however. It is merely a classification variable in these experiments. To attribute causal status to SES in the determination of mental abilities would be to prejudge the issue. SES differences in mental abilities most probably involve both genetic and environmental factors, but this is not at issue in the present study, which aims only to determine the relationship between rote learning and intelligence within low and middle SES groups. According to our theory as presently formulated, the findings with respect to SES differences should be essentially the same regardless of race, although it should be noted that in selecting groups of children who are high or low on SES, our samples represent different proportions of each racial population.

Middle SES Group. In this group the modal occupation of the head of the household was "professional and managerial." All were white. None were welfare recipients.

## Tests and Procedures

Peabody Picture Vocabulary Test. The PPVT is one of the most widely used tests in Headstart programs. It consists of 150 sets of 4 pictures in each; the examiner names one of the pictures in each set and the child is asked to point to the appropriate picture. The PPVT was administered individually to all children by two female examiners, who also administered the other tests in this battery.

Paired-Associates Tests. The paired associates (PA) learning tests were devised by Rohwer (Rohwer, 1967). They consist of 20 pictorial PAs presented by a motion picture projector. There are two conditions of filmed presentation. In one, the picture pairs are motionless objects shown as two separate pictures side-by-side. In the other, the two pictured objects are shown in motion, involving some meaningful action sequence (e.g., a DOG walking to a GATE and closing it). The two conditions are henceforth referred to as Still vs. Action. The PA pictures presented are accompanied by the examiner's verbalization, which took two forms: Names vs. Sentences. In the former, as each PA was presented, E uttered aloud the names of the two objects in view; in the latter, E uttered a sentence containing the two names and relating them in some meaningful action. Thus the 20-item PA tasks include the four conditions: Name-Still, Name-Action, Sentence-Still, Sentence-Action. The test thus yields four scores, one for each condition.

Each S was asked to learn the 20-item PA list by the pairing test method. Two complete trials, two pairing and two test, were administered to each S. Both the visual and auditory materials were recorded on video tape so that the presentations were uniform for all Ss.

Each S was tested individually. After entering the testing room, he was seated in front of the video monitor which was placed at eye level. The examiner (white female) read the instructions, telling the S he was to learn a list of pairs such that when presented with one of the objects from each pair he could recall the other. Immediately after these instructions were given, a practice test was administered, consisting of one pairing and one test trial on four sample pairs. If the S did not respond or responded inappropriately during the test trial of the practice test the list was presented again to insure that the instructions had been understood. The practice test was followed by the presentation of the first pairing trial of the 20-item PA list.

During each of the two pairing trials the 20 PAs were presented at a 4-sec. rate. The two objects in a pair appeared on the screen and simultaneously the verbalization was presented through a speaker. There was a 4-sec. interval between the pairing and test trials. Following this interval, one of the objects from each of the 20 pairs was presented at a 10-sec. rate. The stimulus member of the pair was visible on the screen for only 4 secs., however, such that there was, in effect, a 6-sec. intertrial interval. As the stimulus member of a pair appeared on the screen, its name was presented over the speaker and the S was told to say aloud the name of the object that had appeared with it on the pairing trial. This procedure was repeated for a total of two complete trials.

Serial Learning Test. This test consisted of a set of ten cards, one blank and each of the rest bearing a colored picture of a familiar object: (blank), clock, toaster, fish, crescent moon, saw, ear of corn, mop, telephone, bird. The serial order of presentation (as listed above) was the same for all Ss. The test was administered individually and never on the same day as any other test for any given S.

The E begins by saying, "I have some cards here with pictures on them. We're going to play a game where you try to remember what picture was first, then what picture was next, and so on. First, let's go through the pictures and you tell me what each picture is." (E starts stopwatch.) "What's this one?. . . Right, a clock," and so on through the series. If the child does not know the name, E provides it and records this fact. If the child uses a label other than the usual one for that picture, E allows that name as the accepted one and writes it on the answer sheet. A maximum of 30 seconds is allowed for naming each picture on this naming trial. E notes total time for naming.

After all nine pictures have been thus named, the test trials begin. E says, "All right, now let's go back to the beginning. What was the first picture you saw?" A maximum of 30 seconds is allowed for the child to anticipate the name of each picture, whereupon it is revealed and he is asked, "What is the next picture?" and so on, for a total of 10 trials. At the end of each trial, E says, "Now let's go back and see how many you can remember this time." At the end of the tenth trial the time is recorded.

Ten scores are derived from the S's responses. These are described in the section on results. The single most important score for our purposes is the total number of correct anticipations made by S in the ten learning trials. The maximum possible score is 10 x 9 = 90.

Digit Span Test. This test is patterned after the forward digit span tests of the Stanford-Binet and the Wechsler Intelligence Scale for Children. It uses the same digit series from these tests. However, all digit series were administered in order of increasing difficulty, beginning with 2-digit series and going up to 9 digits. All the series were presented to every child and the responses were recorded.

In beginning this test, E says, "I am going to say some numbers and when I am through I want you to say them just the way I do. Listen carefully, and get them just right. Listen; say 3-5." (E gives further examples if necessary: 1-6, 2-9.) Now say. . ." and E begins with the 2-digit series. Before each digit series E repeats, "Listen carefully, and get them just right." The digits are read at a 1-sec. rate. The series are as follows in order of presentation; the Stanford-Binet form always preceded the WISC at each series length.

| Stanford-Binet | WISC |
|---|---|
| 4-7 | |
| 6-3 | |
| 5-8 | |
| 6-4-1 | 3-8-6 |
| 3-5-2 | 6-1-2 |
| 8-3-7 | |
| | 3-4-1-7 |
| | 6-1-5-8 |
| 3-1-8-5-9 | 8-4-2-3-9 |
| 4-8-3-7-2 | 5-2-1-8-6 |
| 9-6-1-8-3 | |
| 4-7-3-8-5-9 | 3-8-9-1-7-4 |
| 5-2-9-7-4-6 | 7-9-6-4-8-3 |
| 7-2-8-3-9-4 | |
| | 5-1-7-4-2-3-8 |
| | 9-8-5-2-1-6-3 |
| | 5-3-8-7-1-2-4-6-9 |
| | 5-2-6-9-1-7-8-3-5 |

A number of scores were derived from S's responses; these are described in the next section.

# Results

## Chronological Age

The means and SDs of the chronological age (in months) of the middle and low SES groups are 50.15, SD = 7.40 and 52.14, SD = 6.14, respectively -- a mean difference of 2 months in favor of the low SES group.

## Mental Age and IQ

For PPVT Mental Age (in months), the low SES group has $\bar{X}$ = 48.41, SD = 22.55 and the middle SES group has $\bar{X}$ = 64.46, SD = 19.06, a difference significant beyond the .01 level.

The IQs are: low SES $\bar{X}$ = 91.18, SD = 20.13 and middle SES $\bar{X}$ = 109.71, SD = 16.14, a difference significant beyond the .01 level.

Paired Associates Test

Reliability. The reliability of the total score on the PA test, as determined from the intraclass correlation between trials 1 and 2, is .90 for the low SES group and .91 for the middle SES group. Thus, this test has a reliability in both lower and middle class population samples that compares favorably with the reliabilities of the best individual intelligence tests such as the Stanford-Binet and the Wechsler Intelligence Scale for Children.

SES Difference. An analysis of variance of these data (Trials X SES Group) shows that the total score means differ significantly at the .01 level. The means and SDs for the low and middle SES groups are 12.01, SD = 7.48, and 16.60, SD = 7.90, respectively. Thus, the hypothesis of no difference between low and middle SES groups is not borne out by this PA test. The explanation could be either that the hypothesis is wrong or that this particular PA test is administered in such a way as to involve Level II cognitive abilities to a larger extent than is characteristic of other rote learning tasks. The "naming" condition and especially the "sentence" condition of administering the PA test may well make it less "rote" and thus less Level I in nature than would be true for PA tests that are unaccompanied by verbalizations that prompt Level II-type mediational processes.

Although the SES difference is fully significant, it should be compared with the SES difference in IQ in sigma ($\sigma$) units of the middle SES group. The low and middle SES groups differ by 1.15$\sigma$ in IQ; they differ by 0.63$\sigma$ on the PA test. Thus they differ 1.8 times (1.15/.63) as much on IQ as on the PA test. So our hypothesis is partially confirmed: the middle and low SES groups differ much less in PA learning than in IQ.

Correlations Between PA, MA, and IQ. Table 1 shows the intercorrelations between the intelligence test and PA test variables for

- - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - -

the two SES groups. The important correlations from the standpoint of our theory are those between MA and the PA Tasks and between IQ and the PA Tasks. For MA, the correlation with the total PA score is .58 for the middle SES and .20 for the low SES group. When chronological age is partialed out, the correlations between MA and PA learning are .51 and .10 for the middle and low SES groups, respectively. The IQs show a similar difference. These correlations are fully in accord with our hypothesis that Level I (e.g., PA learning) and Level II (e.g., intelligence test scores) measures are more highly correlated in the middle SES than in the low SES population.

Table 1

Intercorrelations among Eight Variables for the Middle SES

and Low SES Samples ($N$ = 100 in Each Group)

Middle SES

| | CA | MA | IQ | NS | NA | SS | SA |
|---|---|---|---|---|---|---|---|
| Chronological Age | -- | | | | | | |
| PPVT Mental Age | .41** | -- | | | | | |
| PPVT IQ | -.01 | .81** | -- | | | | |
| Naming Still | .38** | .53** | .39** | -- | | | |
| Naming Action | .37** | .43** | .31** | .48** | -- | | |
| Sentence Still | .14 | .43** | .41** | .55** | .55** | -- | |
| Sentence Action | .32** | .52** | .42** | .61** | .54** | .58** | -- |
| Total PA | .36** | .58** | .47** | .81** | .78** | .83** | .85** |

Low SES

| | CA | MA | IQ | NS | NA | SS | SA |
|---|---|---|---|---|---|---|---|
| Chronological Age | -- | | | | | | |
| PPVT Mental Age | .26** | -- | | | | | |
| PPVT IQ | -.01 | .81** | -- | | | | |
| Naming Still | .25* | | .03 | -- | | | |
| Naming Action | .37** | .22* | .19 | .36** | -- | | |
| Sentence Still | .39** | .26** | .26** | .32** | .55** | -- | |
| Sentence Action | .28** | .14 | .15 | .44** | .71** | .48** | -- |
| Total PA | .41** | .20* | .21* | .66** | .84** | .76** | .85** |

* $p < .05$

** $p < .01$

## Serial Learning

Reliability. Since only one form of the serial learning test was used, there is no satisfactory method for determining the test's reliability.

Factor Analysis of Serial Scores. Since a number of measures can be derived from the serial learning data, and since there is no good *a priori* basis for selecting one measure over another, it was decided to score the *S*'s protocols on a number of measures and to subject these to a factor analysis (varimax rotation of the principal components) in order to determine the dimensionality of the several measures. The factor analysis was carried out separately in the middle and low SES groups, since the factorial nature of the scores might well be different in lower and middle class populations. Ten measures were entered into the factor analysis:

1. PPVT Mental Age (MA) in months.
2. Number of pictures given no name during the initial naming trial.
3. Number of pictures given an incorrect name during initial naming trial.
4. Number of full trials completed.
5. Number of correct anticipations on the last trial.
6. Number of correct anticipations on the best trial.
7. Total number of overt errors (not including omissions) on all trials.
8. Total number of overt verbal responses (whether correct or incorrect).
9. Study time (sec.) on initial naming trial.
10. Total test time (sec.) for all learning trials.

Total number of correct responses was not included in this factor analysis because it is simply a linear function of two other variables: number of verbal responses - number of overt errors = number of correct responses. The number of omissions (i.e., failures to respond) is not included because it is a linear function of total possible score - number of verbal responses. The inclusion of number correct and omissions would therefore not add anything to determining the factorial structure of this set of measures, since they are completely determined by other measures in the set.

Table 2 shows the means and standard deviations of the middle and low SES groups on these 10 variables.

---

Insert Table 2 about here

---

The factor analysis of these scores leads to the conclusion that the most representative of serial learning ability is the total number of correct responses, since in both SES groups Verbal Responses (Variable 7) and Errors (Variable 8) had almost equally high loadings

Table 2

Means and Standard Deviations of Social Learning Measures

for Low and Middle SES Groups ($\underline{N}$ = 100 in each group)

| Variable | Low SES M | Low SES SD | Middle SES M | Middle SES SD | t |
|---|---|---|---|---|---|
| 1. MA (mos.) | 48.41 | 22.67 | 64.46 | 19.16 | -5.40** |
| 2. No Name | 1.14 | 1.27 | 0.67 | 1.17 | 2.72** |
| 3. Incorrect Name | 1.47 | 1.11 | 0.82 | 0.96 | 4.48** |
| 4. No. Trials | 8.79 | 2.24 | 8.28 | 2.33 | 1.57 |
| 5. No. Corr. last trial | 4.44 | 2.64 | 5.43 | 3.08 | -2.44* |
| 6. No. Corr. best trial | 5.27 | 2.37 | 6.03 | 2.67 | -2.13* |
| 7. Errors | 37.07 | 17.21 | 24.91 | 15.67 | 5.22** |
| 8. Overt Response | 68.37 | 25.55 | 60.14 | 25.63 | 2.28* |
| 9. Study Time (sec) | 54.83 | 20.76 | 49.68 | 21.63 | 1.72 |
| 10. Test time (sec) | 944.95 | 298.95 | 786.62 | 285.60 | 3.83** |
| 11. No. Correct | 31.30 | 18.27 | 35.23 | 22.78 | -1.35 |

* $p < .05$

** $p < .01$

on the one factor most clearly interpretable as serial learning ability, and total number of correct responses is a composite function of Total Responses - Total Errors.

The varimax rotated factors obtained from the intercorrelations of the nine serial scores, along with PPVT MA, are shown in Table 3.

- - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - -

The factors are not all equally interpretable, and only the first three factors are at all comparable in the low and middle SES groups.

Factor I. Verbalization and peak performance (+). This factor is difficult to label or interpret, but it seems to be the same in both SES groups.

Factor II. Serial learning ability (-). This is the clearest factor; it reflects (negatively) what is meant by serial learning ability, having its largest loading on total errors. Note that verbal responsiveness (regardless of correctness) is divided between Factors I and II, because it consists both of correct and incorrect responses. Thus verbal responsiveness _per se_ has both positive and negative effects in serial learning.

Factor III. Willingness to name pictures (+). This factor is very clear. _S_s who use more study time in the initial naming trial also fail to name more pictures. Note that this factor correlates (.52) substantially with MA only in the middle SES group.

Factor IV. Ability to name pictures correctly (+). This is significantly correlated with MA only in the middle SES group.

Factor V. This is completely different in the low and middle SES groups. In the low SES group it is nothing but PPVT mental age. Note that no other factor has an appreciable loading on mental age for the low SES group, which means that all aspects of the serial performances are unrelated to MA in this low SES group. There are substantial correlations, however, between MA and certain aspects of the serial performance (Factors III and IV) in the middle SES group.

Factor II most clearly represents serial learning ability. Total overt responses and total overt errors have the highest loadings, and the difference between these two measures is the number correct. Thus the number of correct responses would have at least as high a loading as errors in Factor II, but it is the single most preferable means of learning ability on this task, since the overt error score does not reflect omissions. In a serial test in which all _S_s are required to learn to a common criterion, rather than being given a constant number of trials as in the present experiment, the best measure of learning ability is not number correct but number of overt errors plus omissions. In the present study, with a constant number of trials for all _S_s, number correct is perfectly correlated (negatively) with errors +

Table 3

Rotated Factors for Serial Learning Measures and PPVT Mental Age

in Low (L) and Middle (M) SES Groups($\underline{N}$ = 100 in each group)

| Variable | I | | II | | III | | IV | | V | |
|---|---|---|---|---|---|---|---|---|---|---|
| | L | M | L | M | L | M | L | M | L | M |
| 1. MA | .10 | .33 | .10 | -.06 | .20 | .52 | -.01 | .50 | -.94 | .08 |
| 2. No Name | -.18 | -.13 | -.21 | .04 | -.88 | -.94 | .11 | .09 | -.09 | -.03 |
| 3. Incorrect Name | -.18 | -.02 | .14 | .13 | -.08 | -.02 | -.85 | -.95 | -.10 | -.06 |
| 4. No. Trials | .47 | .62 | .79 | .57 | .17 | .09 | .24 | -.05 | .02 | -.44 |
| 5. No. Correct Last Trial | .96 | .94 | .05 | -.03 | .11 | .19 | .01 | .10 | .06 | .16 |
| 6. No. Correct Best Trial | .92 | .94 | .16 | -.01 | .19 | .21 | .09 | .12 | .09 | .10 |
| 7. Errors | -.12 | .15 | .92 | .94 | .04 | -.03 | -.09 | -.11 | .06 | -.10 |
| 8. Overt Responses | .51 | .65 | .80 | .73 | .16 | .11 | -.05 | -.05 | .13 | -.04 |
| 9. Study Time | -.11 | -.22 | .01 | -.21 | -.91 | -.69 | -.15 | -.43 | -.16 | -.18 |
| 10. Test Time | -.19 | -.16 | .44 | .13 | -.10 | -.13 | .61 | -.09 | -.27 | -.95 |

omissions. Subsequent discussions of serial learning ability are based on total number of correct responses.

Correlations of Serial Learning and MA. The correlation between number of correct responses in serial learning and PPVT MA is .27 for low SES and .49 for middle SES; with chronological age partialed out, these correlations become .10 and .36, respectively. As in PA learning, there is also a higher correlation between serial learning and intelligence for the middle SES than for the low SES group.

SES Difference. The mean number of correct responses in serial learning is shown as Variable 11 in Table 2. Note that although the PPVT Mental Age means of low and middle SES groups differ by 0.84σ, the serial learning means (number correct) differ by only 0.17σ (σ units based on middle SES group). The SES groups thus differ almost five times as much on PPVT as on serial learning. This clearly supports the hypothesis that low and middle SES groups differ less on rote learning (Level I) tests than on measures of intelligence (Level II). The question of whether the explanation lies in the "culture loading" of the PPVT is taken up in a later section of this report. It will be shown that the PPVT discriminates less between Negro and white groups (who also differ in SES) than a less culture loaded test of intelligence (Raven's Progressive Matrices).

Memory Span Test

Reliability. We have no direct reliability measurement on the digit span tests used in this study, but some idea of their reliability may be gained from the correlation between the Stanford-Binet and WISC digit span. The correlation between the two for the low SES group is .49 and for the middle SES group it is .62. Only two measures enter into this correlation: the longest series gotten right on the Stanford-Binet vs. the longest series gotten right on the WISC. These correlations, then, represent the reliability of digit span based on a single measure. The reliabilities of the average of the two spans would be .65 for low SES and .76 for middle SES.

SES Difference in Digit Span. Table 4 shows the mean digit span for the low and middle (labeled Hi) SES groups. The average of the

- - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - -

Binet and WISC is 3.86 for low SES and 3.88 for middle SES. The means of the total digits given in the correct position for all series are 43.65, SD = 14.83 for low SES and 43.24, SD = 16.32 for middle SES. These mean differences are negligible, as are the differences in SDs. Table 4 also shows the mean number of digits correct for each length of digit series from 2 to 9 digits. These separate digit series have been scored in two ways: (a) Position (Pos.) -- the number of digits recalled in the correct absolute position, and (b) Sequence (Seq.) -- the number of digits correct in adjacent sequence, regardless of absolute position. (Since the maximum possible sequence score is always

Table 4

Means, Standard Deviations, and Correlations with Intelligence Factor in Low and High Socioeconomic Groups ($\underline{N}$ = 100 in each group)

| Variable | | Mean | | | | Standard Deviation | | | | Factor Loadings | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Lo-SES | | Hi-SES | | Lo-SES | | Hi-SES | | Lo-SES | | Hi-SES | |
| Mental Age (mos.) | | 48.41 | | 64.46 | | 22.67 | | 19.16 | | .504 | | .512 | |
| Binet Digit Span | | 3.72 | | 3.63 | | 1.05 | | 1.07 | | .047 | | .482 | |
| WISC Digit Span | | 3.99 | | 4.12 | | 1.02 | | 1.12 | | .073 | | .613 | |
| | | Pos. | Seq. | Pos. | Seq. | Pos. | Seq. | Pos. | Seq. | Pos. | Seq. | Pos. | Seq. |
| Digit Series | 2 | 1.99 | 1.99 | 1.99 | 1.99 | .05 | .05 | .09 | .05 | .032 | .032 | .023 | .023 |
| | 3 | 2.82 | 2.85 | 2.88 | 2.91 | .40 | .31 | .38 | .29 | .138 | .181 | .214 | .210 |
| | 4 | 3.06 | 3.20 | 3.02 | 3.13 | 1.13 | .88 | 1.15 | .95 | .023 | .010 | <u>.877</u> | <u>.870</u> |
| | 5 | 2.00 | 2.46 | 1.83 | 2.42 | 1.32 | .98 | 1.58 | 1.21 | .157 | .156 | <u>.563</u> | <u>.511</u> |
| | 6 | 1.02 | 2.01 | 1.05 | 1.95 | 1.03 | .83 | 1.03 | .90 | .340 | .478 | .372 | .273 |
| | 7 | .54 | 1.53 | .56 | 1.63 | .65 | .63 | .84 | .88 | .325 | <u>.534</u> | .072 | .017 |
| | 8 | .41 | 1.66 | .38 | 1.46 | .49 | .71 | .60 | .65 | .138 | <u>.698</u> | .057 | .026 |
| | 9 | .26 | 1.71 | .28 | 1.71 | .37 | .83 | .49 | .91 | .148 | <u>.760</u> | .133 | .194 |

one less than the maximum possible position score, we have added 1 to the sequence score in every case, to make it directly comparable to the position score.) As one can see in Table 4, there is no appreciable SES difference for any series length for either the position or the sequence scores, and the same is true for their SDs.

Thus it appears that digit memory shows less SES difference than any of the other tests in the battery. It is probably the purest measure we have of what is meant by Level I ability. This is especially interesting in view of the fact that in the general population, composed mostly of the middle class, the digit span test is a quite good measure of intelligence. The reason they have often been regarded as being poor measures of intelligence is that in the brief form in which they are given as part of standard tests such as the Stanford-Binet and the Wechsler tests, the digit span subtest has relatively low reliability and therefore does not display as high correlations with the total IQ as do some of the more reliable subtests such as vocabulary and block design. Yet as early as two and one-half years of age the digit span test correlates .75 (corrected for attenuation) with Stanford-Binet IQ in the normative population (Terman & Merrill, 1960). Digit span also correlates .75 (corrected for attenuation) with adult Wechsler IQ in the normative population, and in a factor analysis of the subscales of the Wechsler Adult Intelligence Scale digit span has a correlation of .80 with the general intelligence factor common to all the subtests (Wechsler, 1958). Since the Wechsler digit span measure is a composite score of memory span for digits forward and digits backwards (i.e., recalled in reverse of the order of presentation), it probably correlates somewhat more with IQ than would just digits forward. Since digits backwards requires some transformation of the input prior to recall, it probably involves some degree of Level II functioning, which would cause it to correlate more with total IQ. Horn (1970) has reported higher g loading for backward than for forward digit span.

A factor analysis of all the variables in Table 4 was carried out separately in the low SES and middle SES groups in order to identify the one factor most clearly identifiable as intelligence. The PPVT MA showed a significant loading on only one factor, in both low and middle SES groups, which is therefore identified as an intelligence factor. The loadings of MA and of the memory span scores on this intelligence factor are shown in the last two columns of Table 4. First of all, note that Binet and WISC digit spans have substantial loadings on this factor in the high SES group and practically zero loadings in the low SES group. Also note that on the position scores of the separate digit series the low SES group shows no substantial loadings, while the high SES group shows very substantial loadings on those series lengths (4 and 5)that are close to or barely exceed the Ss' average digit span. Note, however, that the low SES group shows significant loadings on the intelligence factor on digit series that greatly exceed their memory span and only for sequence scoring. We know that when the number of digits presented exceeds the S's memory span (i.e., the longest series he can recall after a single presentation), he resorts to a simpler strategy of merely associating adjacent digits with little regard for absolute position or other more complex organizing relationships within the series. This change in the encoding process has been found in university students when presented with supraspan series of 12 to 15 digits (Jensen, 1965). This particular form of associative learning appears to be the only component of the low SES group's digit recall performance that has any significant correlation with their intelligence test performance, and since this component has no appreciable relationship with the intelligence factor in the high-SES group, it suggests that the intelligence test (PPVT) itself may be measuring somewhat different mental processes in the two SES groups. Table 5 shows the correlations between position and sequence scores in the high and

- - - - - - - - - - - - -
Insert Table 5 about here
- - - - - - - - - - - - -

low SES groups. Note that the correlations diminish rapidly in the series that exceed the Ss' average memory span, and that the decrease is much more pronounced in the low SES group. The SES differences in correlations for series lengths 7, 8, and 9 are all significant beyond the .05 level. This suggests that the high SES group recalls the series in a global way, so that the position and sequence scores will be highly correlated; if S cannot recall position, he also cannot recall sequence. There seems to be an organization of the input into a total gestalt, such that memory failure affects every aspect of the gestalt -- position and sequence alike. In the low SES group, on the other hand, there is greater dissociation between position and sequential knowledge. Since for short series, one type of encoding or the other yields much the same results and so the correlation between position and sequence scores is high. But when the series are long (7, 8, 9), the dissociation between the position and sequential associative memory can show up. We may characterize the high SES group as acquiring an overall picture or organized gestalt of the whole series, the memory trace of which is subject to more or less uniform and global decay. The low SES Ss, on the other hand, seem to learn "what is next to what" in the series, and these adjacent sequential associations seem to be retained independently of position information. At the very least, the correlational differences shown in Tables 4 and 5 suggest differences between the SES groups in the processes involved in memory span performance and its relationship to the intelligence components measured by the PPVT.

## Intercorrelations Among the Major Variables

Table 6 shows the intercorrelations among the major variables described in previous sections. In the case of PA, serial learning, and digit span, the correlations are based on the total number of correct responses for the entire test. Also shown are the means, SDs,

- - - - - - - - - - - - -
Insert Table 6 about here
- - - - - - - - - - - - -

the mean differences in middle SES sigma units, and the t test for

Table 5

Correlation Between Position and Sequence Scoring

of Digit Series Test

| | Series Length | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SES | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| High | 1.00 | .98 | .93 | .93 | .85 | .60 | .47 | .39 |
| Low | 1.00 | .95 | .91 | .90 | .83 | .29 | .16 | -.01 |

Table 6

Intercorrelations, Means, SDs, and SES Differences among Major Variables in Low (Below Diagonal), and Middle (Above Diagonal) SES Groups

(N = 100 in each group)

| Variable | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 1. Age (Mos.) | | | .41 | -.01 | .36 | .51 | .46 |
| 2. MA (PPVT) | | .26 | | .81 | .58 | .49 | .39 |
| 3. IQ (PPVT) | | -.01 | .81 | | .47 | .30 | .27 |
| 4. Paired Associate | | .41 | .20 | .21 | | .55 | .52 |
| 5. Serial | | .32 | .27 | .26 | .37 | | .50 |
| 6. Digit Memory | | .16 | .34 | .28 | .26 | .45 | |
| Mean | Middle SES | 50.15 | 64.46 | 109.71 | 16.60 | 35.23 | 43.20 |
| | Low SES | 52.14 | 48.41 | 91.10 | 12.01 | 31.30 | 43.65 |
| SD | Middle SES | 7.40 | 19.06 | 16.14 | 7.90 | 22.78 | 16.32 |
| | Low SES | 6.14 | 22.55 | 20.13 | 7.48 | 18.27 | 14.83 |
| Mean Difference/$SD_M$ | | -.269 | .842 | 1.15 | .581 | .173 | -.028 |
| t | | -2.07 | 5.40 | 7.21 | 4.25 | 1.35 | -.20 |
| | | * | ** | ** | ** | | |

* p < .05

** p < .01

the significance of the difference. The average of the correlations in Table 6 is .44 for the middle SES and .31 for the low SES group. More important is the comparison of the average correlation among PA, Serial, and Digit Memory (.52 for middle SES; .36 for low SES) with the average correlation between MA and the Level I (PA, Serial Digits) measures (.49 for middle SES; .27 for low SES). If the among correlations can be treated like a reliability coefficient for the Level I measures, the average correlation between MA and Level I measures can be, in effect, corrected for attenuation by dividing them by the "among" correlations. Thus "corrected," they are .94 for middle SES and .75 for low SES.

Multiple Correlation. Probably the best way of comparing the SES groups on the overall correlation between the learning and memory test, on the one hand, and the intelligence test, on the other, is by means of the multiple-R. Fourteen variables were used to predict MA. The 14 variables were:

1. Chronological Age
2. Serial Learning (Total Score)
3. PA Learning (Naming-Still)
4. PA Learning (Naming-Action)
5. PA Learning (Sentence-Still)
6. PA Learning (Sentence-Action)

7-14. Digit series of 2 to 9 digits (number in correct position)

The multiple correlation, R, between the 14 predictor variables and PPVT MA is 0.54 for the low SES and 0.71 for the middle SES group. The difference is significant at the .05 level. In terms of proportion of variance in MA predicted by the 14 variables, indicated by $R^2$, the corresponding values are 0.29 (for low SES) and 0.51 (for middle SES).

These results thus are consistent with the hypothesis of a higher degree of relationship between associative learning abilities (Level I) and intelligence (Level II) in middle SES Ss than in low SES Ss.

## Summary

Low and middle SES preschool children were compared on Peabody Picture Vocabulary (MA and IQ) as a measure of cognitive ability and on tests of paired associate learning, serial learning, and digit memory span as measures of associative learning ability. The SES groups differed much more on the intelligence measures (MA and IQ) than on any of the learning tasks, and they differed virtually not at all in memory span. Correlations among all tasks were generally higher for the middle SES than the low SES groups, and the middle SES group showed a consistently higher relationship between the intelligence and learning measures than did the low SES group. The results, both with respect to mean SES differences and to the correlation between intelligence and rote learning measures are consistent with the hypothesis of an interaction between SES, intelligence, and learning ability, as formulated in the introductory part of this report.

Level I and Level II Performance in Low and Middle

SES Elementary School Children

Arthur R. Jensen

In the preschool study digit memory span showed the least difference between the low and middle SES groups. It also showed differential correlations with the Level II (intelligence) factor in the low and middle SES groups. For these reasons, and because digit span performance corresponds closely to our theoretical conception of Level I ability -- to register, retain, and recall stimulus inputs -- it was decided to investigate the relationship between digit memory and intelligence in older school children in Grades 4, 5, and 6, since by that age intelligence test scores have become relatively stable and representative of intellectual measures obtained at subsequent ages up to adulthood (Bloom, 1964).

The principal questions are: (1) Is the SES difference smaller in the memory tests than in the intelligence test, as in the preschool study? and (2) Is the correlation between digit memory and intelligence lower among low SES than among middle SES children?

These questions were investigated in three studies.

## Study I

In the first study the aim was to compare the Level II ability (abstract intelligence) of groups of children who were selected for being either very high or very low in Level I ability as indexed by digit memory, and to make this comparison within low and middle SES groups. The theory predicts that the low SES group will show a smaller difference and also a lower correlation between the Level I and Level II measures.

### Method

#### Subjects

Ss were drawn from two highly contrasting schools in the East Bay Area of San Francisco. These schools were selected because, in one case, the student population is typical of children who can be characterized as of low SES. Their average level of scholastic performance is considerably below the average of national norms. The general mean IQ (Lorge-Thorndike) of the elementary school was 85, SD = 12. Nearly all the pupils were Negro and nearly all came from neighborhoods which can be classed as of lower SES. The pupils of this school are thus quite typical of those for whom programs such as Headstart are particularly intended. The contrasting school was in a middle or upper-middle class white neighborhood where the majority of heads of household are employed in managerial and professional occupations. The overall level of scholastic performance in the school is well above national norms and the school's mean IQ

(Lorge-Thorndike) was 113, SD = 13. Thus the Ss in this study were drawn from schools that had about an equal but opposite deviation from the general population mean of 100 in IQ -- resulting in a total mean difference between the low and middle SES groups of approximately 2 standard deviations.

The numbers in each school, labeled low SES and middle SES, were distributed as follows:

| Grade | Low SES | Middle SES |
|---|---|---|
| 4 | 141 | 175 |
| 5 | 123 | 150 |
| 6 | 117 | 164 |
| Total | 381 | 489 |

## Tests

Memory for Numbers. This test was the measure of Level I. A more elaborate and reliable digit memory test was desired than the one used in the preschool study, which consisted of a combination of the digit span subtests from the Stanford-Binet and the Wechsler Intelligence Scale for Children. Therefore a new short-term memory test was specially devised for this study. It was an auditory memory test. The entire test, including the instructions to the subject, were tape recorded (by a clear male voice) to insure uniformity of presentation. The test as it was read into the tape recorder (except for the headings and timing indications) is given in Appendix A.

The test has three parts, each preceded by a short practice test, and each consisting of three series of three digits. The practice test allows the S to become familiar with the procedure of the tests that follow it. If an S fails the practice tests it can usually be assumed he has not understood the instructions or for some reason is not cooperating. It is rare when a normal child beyond second grade misses any of the practice tests. In each part, E utters a series of from 4 to 9 digits. There are three replications or equivalent forms of each length of series. The digits are read by E at precisely a 1-sec. rate this was achieved by recording (on a dictaphone) a metronome ticking at a 1-sec. rate and having E listen to it through an ear phone (on one ear) while reading the numbers aloud. Each digit series was followed by the sound of a bong, which was the signal for S to write as many digits in corrected order as he could recall. Specially prepared answer sheets were provided (see Appendix B).

Part I is immediate recall (I). After a single presentation of the series, the bong sounds immediately after the last digit and the S writes his answer at once. After 13 seconds for writing, the bong signals the S to pay attention for the next series.

Part II is repeated series (R). It is like Part I except that each series is given three times in succession, separated by a 1-sec. interval filled by a low rumbling tone (called "noise" in the instructions). Thus Part II is not only a measure of short-term memory span but of learning as well, since S hears the same series three times in succession.

Part III is delayed recall (D). In this condition the bong does not follow until 10 seconds of silence have elapsed since the last digit in the series. Ss are instructed to hold up their pencils until the bong is sounded. The 10 seconds delay interval permits S to rehearse to himself or, in the absence of rehearsal, allows time for some delay in the memory trace. Earlier studies indicate that a 10 sec. delay almost invariably results in some retention loss in digit memory in college students, although in these earlier studies the delay interval was filled with mildly distracting stimuli, and so they may not be comparable to the present test (Jensen, 1965). Since we wished to measure individual differences rather than test the effects of an experimental variable, we did not counterbalance the order of presentation of the three parts in the same order, viz., I. R, D.

The test was administered to intact classrooms. While the tape recorder was being played from the front of the room, E and an assistant assumed positions in the room from which they could observe whether children were following directions. This was facilitated by using three colors of paper in the test booklets: children on the wrong page could be quickly spotted.

Scoring. The S's score on each part is the total number of digits recalled in the correct position over all series. "Correct position" is unambiguously identified by the "boxes" on the answer sheets.

## Results

### SES Mean Difference

Table 7 shows the mean number of correct responses for each of the Memory for Number subtests in the low SES and middle SES groups.

Insert Table 7 about here

The last column in Table 7 shows the ratio of the mean difference between low and middle SES groups to the SD of the middle SES group. This ratio is the most meaningful means of comparing the SES groups. Because of the large sample sizes, all of the differences are significant well beyond the .01 level. But more important is the fact that the differences between the SES groups are quite large, averaging 1.29 sigmas. This finding is greatly at variance with the results of the preschool study, in which the mean difference (on a different and individually administered digit memory test)

Table 7

Means, SDs, and Mean SES Difference in Sigma Units (of Middle SES group) for Low SES and Middle SES Groups on Memory for Numbers Tests (I = Immediate Recall, R = Repeated Series, D = Delayed Recall)

| Grade | Test | Mid SES N | Mid SES M | Mid SES SD | Low SES N | Low SES M | Low SES SD | $(\bar{X}_M - \bar{X}_L)/SD_M$ |
|---|---|---|---|---|---|---|---|---|
| 4 | I | | 66.3 | 14.1 | | 47.7 | 13.9 | 1.32 |
| | R | | 71.1 | 16.1 | | 51.7 | 15.2 | 1.20 |
| | D | | 58.9 | 16.4 | | 37.1 | 14.6 | 1.33 |
| | Tot. | 175 | 197.0 | 42.8 | 141 | 136.7 | 38.1 | 1.41 |
| 5 | I | | 72.9 | 14.2 | | 54.7 | 14.3 | 1.28 |
| | R | | 78.2 | 15.5 | | 60.9 | 14.9 | 1.12 |
| | D | | 64.1 | 16.1 | | 47.9 | 14.0 | 1.01 |
| | Tot. | 150 | 215.5 | 39.7 | 123 | 163.6 | 37.7 | 1.31 |
| 6 | I | | 77.6 | 14.7 | | 60.0 | 14.9 | 1.20 |
| | R | | 84.5 | 13.0 | | 67.1 | 17.1 | 1.34 |
| | D | | 71.0 | 13.9 | | 55.9 | 14.3 | 1.09 |
| | Tot. | 164 | 233.1 | 37.1 | 117 | 182.9 | 41.0 | 1.35 |
| Combined | I | | 72.10 | 15.12 | | 53.71 | 15.19 | 1.22 |
| | R | | 77.74 | 15.92 | | 59.40 | 16.96 | 1.15 |
| | D | | 64.52 | 16.33 | | 46.39 | 16.29 | 1.11 |
| | Tot. | 489 | 214.72 | 42.75 | 381 | 159.55 | 43.35 | 1.29 |

between low and middle SES groups was only .03 (in favor of the low SES group). If there is an age trend in the magnitude of the low vs. middle SES difference in digit memory, it is not apparent in the age range from grades 4 through 6. Could the difference be attributable to the fact that this was a group-administered test while the preschool test was administered individually? Are low SES children more easily distracted in a group testing situation? Does their performance require closer supervision by E in order to reach its maximum? Special studies, reported subsequently, were undertaken to help answer some of these questions.

Note also that contrary to what one might expect, the SES groups differ least on the Delayed Recall condition -- the one condition in which covert rehearsal or other verbal mediational processes would be thought to have the greatest effect and consequently give an advantage to the middle SES group. No such advantage appears in these data. The largest SES difference is found for the Immediate Recall condition.

Another interesting point is that the SDs do not differ appreciably in the two SES groups, and the frequency distributions of the scores are relatively normal in both SES groups.

Although the SES groups differ by about 1.3 SDs on the memory test, it should be noted that the schools from which they are a large sample differ on the average about 2 SDs in IQ. So the SES difference for the memory test is only about 1.3/2.0 or 65% as great as for the IQ. If there is a substantial correlation between IQ and memory span in the middle SES population, as postulated by the theory, then we should expect some difference in Level I performance (in this case memory span) between a low and middle SES group when the latter is above the general population mean in IQ (Level II). But the difference of 1.3 sigmas found here seems greater than would have been predicted from the theory, although the theory is not so precisely formulated as yet as to yield exact quantitative predictions. It permits only directional predictions of the "greater than" or "less than" variety in comparing various Level I and Level II tests and their intercorrelations in low and middle SES groups.

Reliability

The reliabilities of the Memory for Numbers Test were determined by means of the intraclass correlations between the three equivalent forms of each subtest. Since a shortened version of the test, made only one-third as long by using only one form of each subtest, was used in a subsequent study, the reliabilities were determined for both the short and the long forms of the test. The reliability coefficients for low and middle SES groups are shown in Table 8.

- - - - - - - - - - - - - -

Insert Table 8 about here

- - - - - - - - - - - - - -

Table 8

Reliability of the Memory for Numbers Test

| Long Form | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Middle SES | | | | Low SES | | | |
| Grade | I | R | D | Tot. | I | R | D | Tot. |
| 4 | .84 | .74 | .88 | .91 | .66 | .86 | .85 | .86 |
| 5 | .76 | .83 | .76 | .81 | .79 | .72 | .71 | .87 |
| 6 | .74 | .86 | .81 | .87 | .83 | .89 | .81 | .86 |
| All Grades | .95 | .95 | .97 | .89 | .91 | .93 | .93 | .89 |

| Short Form | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Middle SES | | | | Low SES | | | |
| Grade | I | R | D | Tot. | I | R | D | Tot. |
| 4 | .64 | .48 | .71 | .77 | .40 | .68 | .66 | .66 |
| 5 | .51 | .63 | .51 | .59 | .56 | .47 | .45 | .69 |
| 6 | .49 | .66 | .59 | .70 | .63 | .73 | .59 | .67 |
| All Grades | .87 | .86 | .91 | .72 | .78 | .81 | .82 | .72 |

The reliabilities of the total scores compare favorably with those for group administered standard intelligence tests, but the reliabilities of subtest scores within grades are not as high. The short form of the test, being only one-third as long, has reliabilities that are below an acceptable level as a basis for individual decisions but are still quite adequate as a basis for group comparisons. The most important point is that the differences in reliability for low and middle SES groups are completely negligible, and for total score over all grades are exactly the same. It can be noted that the reliability of the total score (over all grades) is lower than the reliability of the scores for the individual subtests. This can only mean that the three subtests are somewhat different in factorial composition; that is, they are less highly intercorrelated than are the three equivalent forms of each subtest.

Correlations Among Subtests. Table 9 shows the correlations (over all grades) among the I,R,D subtests before and after correction for attenuation in the middle and low SES groups. Even

Insert Table 9 about here

after correction for attenuation, the subtests have only slightly more than half their true variance ($r_c^2$) in common with one another. The low and middle SES groups do not differ significantly in the pattern of intercorrelations, which suggests that the subtests are measuring similar functions in both SES groups.

## Study II

The aim of the second study is to examine the relationship between Level I and Level II in the low and middle SES groups tested in the preceding study. Since earlier studies had compared below average and above average IQ groups on learning and memory abilities, it was decided to do the opposite in this study, that is, to compare extreme groups in memory span on an intelligence test and to make these comparisons within the low and middle SES groups.

Theoretical Predictions. The theory predicts that low SES children differing markedly in Level I ability (here measured by digit memory) will differ less in Level II ability (here measured by Raven's Progressive Matrices) than will middle SES children. In other words, there should be more high Level I Ss with low Level II performance in the low SES than in the high SES group. A corollary is that there should be a higher correlation between the Level I and Level II measures in the high SES than in the low SES group.

## Method

### Subjects

Ss were drawn from the same 4th, 5th, and 6th grade groups

Table 9

Raw Correlations (r) and Correlations Corrected for Attenuation ($r_c$) Among Immediate, Repeated Series, and Delayed Recall Subtests of the Memory for Numbers Tests in Grades 4, 5, and 6 Combined for Low SES and Middle SES Groups

| Tests | Low SES r | Low SES $r_c$ | Middle SES r | Middle SES $r_c$ |
|---|---|---|---|---|
| I X R | .73 | .79 | .70 | .74 |
| I X D | .67 | .73 | .70 | .72 |
| D X R | .75 | .80 | .75 | .79 |

who were tested in the previous study. Ss who had obtained the ten highest scores and Ss who had obtained the ten lowest scores on the Memory for Numbers Test were selected from each grade in the low and middle SES schools. The percentile scores for the selection cut-off on Memory for Numbers Total Score in selecting the 10 highest and lowest in each grade are shown below:

| | Middle SES | | Low SES | |
|---|---|---|---|---|
| Grade | Lowest | Highest | Lowest | Highest |
| 4 | 7.1 | 9.3 | 5.7 | 9.4 |
| 5 | 8.1 | 9.2 | 6.7 | 9.3 |
| 6 | 8.5 | 9.1 | 6.1 | 9.4 |

Tests

Memory for Numbers. This was the measure of Level I. The 120 selected Ss were retested on the Memory for Numbers Test; this time the test was administered individually to each S. The reasons for testing these Ss a second time on the same test were twofold: first, so that any regression effects as a result of selecting extreme groups would be allowed to occur, and second, to eliminate possible "flukes" from the extreme groups -- in short, it was a form of double screening which considerably increased the reliability of the Level I assessment. We wished to avoid any overlap between the extreme groups on the Memory test, and this was accomplished. On retest the extreme groups proved relatively homogeneous and showed no overlap of total Memory scores, nor was there any overlap of extreme groups across SES groups. That is, middle SES low scorers did not overlap low SES high scorers, etc.

Raven's Progressive Matrices. The measure of Level II were the Colored Progressive Matrices and the Standard Progressive Matrices. These are nonverbal tests of reasoning ability. They were devised to load heavily on the g factor, in the Spearman sense, and on no other ability factors. The g saturation of the tests, according to the test manual, is close to .80 and the test's reliability is close to .90. The test was administered individually (using the test booklet form) according to the instructions in the test's manual. Ss were self paced and were encouraged to respond to every item until they had missed five of the last six successive items. (One out of six is a chance score.) So as to avoid a ceiling effect, both the children's (Colored) and adult's (Standard) forms of the test were used. The Colored Matrices consist of 36 problems. They are graded in difficulty beginning at a level suitable for a mental age of 3 to 4. The solution to the first problems are so easy that virtually all school-age children "catch on." All problems follow the same basic format, that is, selecting the one out of six multiple choice patterns that best completes the blank space in each matrix. The Standard Matrices consist of 60 such problems, but the first 24 problems are very easy and overlap the Colored Matrices. Therefore, in using the Standard form as a continuation of the Colored form, we began with the 25th item and continued until the S again missed 5 out of the last six problems. Any S who got one or more correct answers in the last 12 on the Colored form was continued on the Standard form, since it would be extremely unlikely that anyone missing all of the last 12 items on the Colored form could score better than chance on the Standard form beginning with the 25th item.

## Results

### Memory for Numbers Test

No attempt was made to estimate the reliability of the scores in the select groups. Since Ss were selected for extreme scores, it would be relatively meaningless to obtain reliabilities within the quite homogeneous extreme groups, and the reliability of the differences between extreme groups is properly determined by analysis of variance. The analysis of variance performed on total scores of the individually administered Memory for Numbers Test had three variables: Schools (low vs. middle SES), Grades (4, 5, and 6) and Level of performance -- the extreme lower and upper groups on the selection test (i.e., the first group-administered Memory for Numbers Test). Effects significant beyond the .01 level are Schools (S), Grades (G), Levels (L), and S X L. No other interactions were significant. The main effect for Levels was, of course, predominant ($F = 430.34$ for 1/108 df), since Ss had been selected so as to avoid any overlap between the extreme memory group, within or between schools. Table 10 shows the mean total scores of the four groups on the memory test for all grades combined.

Insert Table 10 about here

### Raven's Progressive Matrices

Table 11 summarizes the group means on the Matrices. The overall

Insert Table 11 about here

mean difference between the low and middle SES groups is 19.8 raw score units or $2.1\sigma$, which corresponds very closely to the mean IQ difference between the entire low and middle SES student populations.

Inspection of Table 11 reveals that the low and middle SES groups, despite their similarity on the memory test, perform very differently on the Matrices. Although there was absolutely no overlap between the low SES Upper Memory group and the middle SES Lower Memory group (their respective means on the memory test were 223 vs. 167, a difference of $1.86\sigma$), the middle SES Lower Memory group still exceeds the low SES Upper Memory group on the Matrices by an average of 8.4 points or $.89\sigma$. In fact, the low SES 6th grade Upper Memory

Table 10

Mean Total Memory Score in Selected Upper and Lower Groups

on Prior Memory Test

| Prior Level | Low SES | Middle SES | Difference |
|---|---|---|---|
| Upper 30 | 223.36 | 301.76 | 78.40** |
| Lower 30 | 129.63 | 167.30 | 37.67 |
| Difference | 93.73** | 134.46** | |

**p <.01 Mean Square Error = 907.56

$\sqrt{MSE}$ = 30.12

Table 11

Mean Raven Progressive Matrices Scores in Lower and Upper Digit Memory Groups

as a Function of Low vs. Middle SES

| | Middle SES | | | | Low SES | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Level on Memory Test | Grade | | | | Grade | | | | | SES |
| | 4 | 5 | 6 | Mean | 4 | 5 | 6 | Mean | Grand Mean | M-L |
| Upper | 52.2 | 49.8 | 54.8 | 52.3 | 23.2 | 31.1 | 31.1 | 28.5 | 40.4 | 23.8 |
| Lower | 34.8 | 34.4 | 41.4 | 36.9 | 18.7 | 22.9 | 21.6 | 21.1 | 29.0 | 15.8 |
| Mean | 43.5 | 42.1 | 48.1 | 44.6 | 21.0 | 27.0 | 26.4 | 24.8 | 34.7 | 19.8 |
| U - L | 17.4 | 15.4 | 13.4 | 15.4 | 4.5 | 8.2 | 9.5 | 7.4 | 11.4 | |

Mean Square Error = 88.3

$\sqrt{MSE} = 9.4$

group falls below the middle SES 4th grade Lower Memory group by 3.7 points or 0.39σ. On Matrices performance SES appears to be a much more potent variable than short term memory.

Table 12 gives the analysis of variance of the Matrices scores.

- - - - - - - - - - - - - - -
Insert Table 12 about here
- - - - - - - - - - - - - - -

The interaction term SL shows that memory level makes a significantly greater difference to Matrices performance in the middle than in the low SES group, as can be seen directly in Table 11.

Table 13 shows the percentages of the total variation (as indicated by the sum of squared deviations) for the digit memory and

- - - - - - - - - - - - - - -
Insert Table 13 about here
- - - - - - - - - - - - - - -

Matrices tests. The largest source of variation was, of course, forced on the memory test by selecting extreme upper and lower scores on a prior test of memory. But on the Matrices, SES becomes the largest source of variance.

By selecting the 30 highest and 30 lowest in memory span performance from the entire 4th, 5th, and 6th grades in the low and middle SES schools (i.e., the upper and lower 6% to 8%), it could be claimed that we selected not only on ability but on motivation and test-taking attitudes as well. This is probably true. The highest scorers in either SES group were probably better motivated than the lowest scorers in either group. If true, this would make even more impressive the comparison between the Matrices scores of the Upper memory in low SES group and the lower memory in middle SES group, whose mean Matrices are 37 vs. 29, respectively -- a difference of 0.89σ in favor of the middle SES group, although these groups differ in memory scores by 1.86σ in favor of the low SES group. Another way of stating this is that out of the 30 Ss in the low SES school who were above the common mean (of both schools) in digit memory, 22 were below the common mean on the Matrices. On the other hand, out of the 30 Ss in the middle SES school who were above the common mean in digit memory, only 2 were below the common mean on the Matrices.

Correlation Between Digit Memory and Matrices

A nonparametric measure of relationship between digit memory and Matrices performance is called for by these data, since the groups were selected originally for extreme scores on memory span and therefore the bivariate normal distribution required for proper interpretation of the Pearson r does not obtain. Although any kind of correlation obtained for these data could not be regarded as representative of population parameters, they can permit a test of our

Table 12

Analysis of Variance of Raven's Progressive Matrices

| Source | df | ms | F | p < |
|---|---|---|---|---|
| SES (S) | 1 | 11761.2 | 133.20 | .01 |
| Grade (G) | 2 | 250.4 | 2.83 | .01 |
| Level of Memory (L) | 1 | 3898.8 | 44.14 | .01 |
| SG | 2 | 167.3 | 1.89 | ns |
| SL | 1 | 480.0 | 5.43 | .025 |
| GL | 2 | 1.8 | < 1 | |
| SGL | 2 | 51.8 | < 1 | |
| Within | 108 | 88.3 | | |

Table 13

Percentages of Total Variation (Sum of Squares)

Attributable to Main Effects

and Interactions for Digit Memory and Progressive Matrices Scores

| Source of Variation | Digit Memory | Prog. Matrices |
|---|---|---|
| Socioeconomic Status (S) | 15.74 | 44.18 |
| Grade (G) | 4.95 | 1.88 |
| Level of Memory (L) | 60.85 | 14.64 |
| SG | 0.14 | 1.26 |
| SL | 1.94 | 1.80 |
| GL | 0.55 | 0.01 |
| SGL | 0.56 | 0.39 |
| Within | 15.27 | 35.84 |
| Total | 100.00 | 100.00 |

hypothesis that Level I and Level II abilities (represented here by memory span and Matrices, respectively) are more highly related in the middle SES than in the low SES group. (We have already noted one indirect test of this hypothesis in the significant SES X Memory Level interaction shown in Table 12.) To get at this relationship more directly, a nonparametric measure of correlation, the phi coefficient, was obtained between digit memory and Matrices within the low and middle SES groups. Each variable was dichotomized at the median (thus forcing equal marginal frequencies) of the respective SES groups. The results are shown in Table 14. The phi coefficients differ significantly beyond the .02 level and beyond

- - - - - - - - - - - - - - - -

Insert Table 14 about here

- - - - - - - - - - - - - - - -

the .01 level for a one-tailed test, which is justified by the fact that the direction of the difference was hypothesized.

## Study III

The third study in this series was intended to test the hypotheses with which the previous studies were concerned, in a total school population. Up to this point the relationship between Level I and Level II measures had been investigated in quite highly selected groups. Since the predicted direction of differences and correlations were largely borne out at a satisfactory level of statistical significance under these relatively small sample conditions, it next becomes necessary to check the theory in an unselected school population to rule out any chance that the relationships observed in the previous studies were in any way artifactual results of selection. To accomplish this, all 4th, 5th, and 6th grade children in a partially integrated public school system with 50% white and 40% Negro pupil population were tested on Level I and Level II tests (10% are Oriental and other ethnic minorities). In addition, certain controls were introduced in the testing procedures to permit evaluation of extraneous factors, not germane to the theory, that might affect test performance. In this study Ss were grouped by race, as listed by the child or his parents in the school records, rather than by SES, although in this community there is a substantial SES differential between the Negro and white population. But there is also some overlap. If anything, classification by race rather than SES should attenuate the results with respect to the theory, but since the white and Negro populations do in general differ socioeconomically, the same predictions should still pertain. Furthermore, the large number of Ss in this study should leave little doubt about the statistical significance of the results.

### Method

#### Subjects

All the 4th, 5th, and 6th grade pupils in 14 elementary schools

Table 14

Contingency Tables and Phi Coefficients for Relationship

between Digit Memory and Progressive Matrices

in Low SES and Middle SES Groups

Low SES

| | | Matrices | | |
|---|---|---|---|---|
| | | Below Md. | Above Md. | |
| Digit | Above Md. | 10 | 20 | 30 |
| Memory | Below Md. | 20 | 10 | 30 |
| | | 30 | 30 | |

phi = .33

Middle SES

| | | Matrices | | |
|---|---|---|---|---|
| | | Below Md. | Above Md. | |
| Digit | Above Md. | 4 | 26 | 30 |
| Memory | Below Md. | 26 | 4 | 30 |
| | | 30 | 30 | |

phi = .73

were tested on three different days on Level I and Level II tests in addition to certain "control" tests More than a thousand children were tested in each grade, but only the data on the white and Negro groups were analyzed in the present study, and intercorrelations between tests, of course, could be obtained only for children who were in class on each day of testing.

## Procedure

*Ss* were tested in their classrooms by six trained testers on our research staff. The classroom teacher acted only as a proctor. Using specially trained testers helped to insure uniformity of procedures and timing on all tests. A testing supervisor on the research staff observed every tester in action on one or more occasions during the testing program as a form of "quality control" for any deviations from the standard procedures. Half the testers were Negro and half were white. All were college students or graduates. Negro and white testers administered tests in equal numbers of predominantly white and predominantly Negro classes. Teachers always remained in the room to assist in passing out and collecting test booklets, pencils, etc.

## Tests

*Lorge-Thorndike Intelligence Test.* This test (Level III, Form B) was the measure of Level II ability. The form used was designed for children in grades 4 through 6. The test has two main parts: Verbal and Nonverbal. It correlates highly with individual tests of intelligence such as the Wechsler and Stanford-Binet, and both parts have a high loading on the *g* (general intelligence) factor. It is one of the most widely used intelligence tests in schools throughout the United States and is State-mandated in Grades 3 and 6 in California's public schools. In the normative population, the L-T test yields an average IQ of 100, SD = 16.

*Memory for Numbers.* This is the Level I measure. It is exactly the same test, administered by tape recorder, as used in the previous study, except that the short form was used in the present study. (The long form consists of three complete replications, by equivalent forms, of the short form.)

*Listening-Attention Test.* This test, which immediately precedes the Memory for Numbers Test, was intended as a control for the latter test. The Listening-Attention (LA) test is also administered by tape recorder in the same male voice that recorded the Memory for Numbers test. High scores on the LA test indicate that the *S* is able to hear and distinguish correctly the numbers spoken by *E* on the tape, and to follow directions, keep pace with the test, and mark his test answer sheet properly. Children who, for whatever reason, cannot or will not do these things are not up to taking the Memory for Numbers test that follows, and their scores cannot be considered valid measures of their Level I ability.

The LA test is quite simple. (The test booklet is shown in Appendix C.) It begins with two short practice series, *a* and *b*.

E says, "Put the point of your pencil on the letter a. Now, I am going to say one number in each pair, and you should cross out the number I say -- cross it out with an X. Ready? 2-4-8-9-3." (The numbers are spoken at a 2-second rate.) The rest of the test proceeds in the same fashion. At the beginning of each series, S is told to put his pencil on the letter at the top of the list. There are 100 items in all; the S's score is the total number correct.

Test of Speed and Persistence. This test, called the Making X's test, is intended as an assessment of test-taking motivation. It was always given just prior to the Lorge-Thorndike Intelligence Test. The Making X's test gives an indication of the S's willingness to comply with instructions in a group testing situation and to mobilize effort in following these instructions for a brief period of time. The test involvea no intellectual component, although it may involve a motor skills factor, especially in young children. Most of the individual differences in scores, however, is probably attributable to Ss' effort and motivation. Children who have already been in school one or more years and are thereby experienced in the use of paper and pencil perform on this test in accord with their willingness to exert effort under instructions to do so. Children (with the exception of those with sensorimotor handicaps) who do very poorly on this test, it can be suspected, are not likely to reflect their true level of ability.

The Making X's test (shown in Appendix D) consists of two parts. On Part I the S is asked simply to make X's in a series of squares for a period of exactly 90 seconds (timed precisely with a stopwatch). In this part the instructions say nothing about speed; they merely instruct the child to make X's. The maximum possible score on Part I is 150, since there are 150 squares provided in which the child can make X's. After a 2-minute rest period the child turns the page of the test booklet to Part II. There the child is instructed to show how much better he can perform than he did on Part I and to work as rapidly as possible. The child is again given 90 seconds to make as many X's as he can in the 150 boxes provided. The gain in score from Part I to Part II reflects both a practice effect and an increase in motivation and effort as a result of the instructions to the S to work as rapidly as possible and exceed his performance on Part I.

## Results

### Control Tests

Listening-Attention. Summary statistics on the LA test are shown in Table 15. As can readily be seen from this table, the level of

- - - - - - - - - - - - - - -
Insert Table 15 about here
- - - - - - - - - - - - - - -

performance is very high on this test. The mean is close to a perfect score in all grades, and even the lower quartile ($Q_1$) is still a perfect score. The median is 100, a perfect score. In short, virtually

Table 15

Statistics on the Listening-Attention Test for White (W) and Negro (N) Groups

| Statistic | Grade 4 | | Grade 5 | | Grade 6 | |
|---|---|---|---|---|---|---|
| | W | N | W | N | W | N |
| N | 504 | 411 | 477 | 416 | 442 | 387 |
| Mean | 98.3 | 98.2 | 99.3 | 98.6 | 99.6 | 99.2 |
| SD | 11.9 | 7.6 | 6.1 | 6.0 | 5.1 | 5.8 |
| $SE_M$ | 0.53 | 0.37 | 0.28 | 0.29 | 0.24 | 0.30 |
| Min. | 0.0 | 0.0 | 0.0 | 41.0 | 0.0 | 0.0 |
| Max. | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Range | 100.0 | 100.0 | 100.0 | 59.0 | 100.0 | 100.0 |
| Median | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $Q_1$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $Q_3$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

all Ss obtained a perfect score on this test, showing that these groups are quite in possession of the prerequisite skills needed for the Memory for Numbers test. That is to say, they can follow the directions and they can hear and discriminate numbers as spoken by the male voice on the tape recorder. Even with the large Ns, there is no significant difference on this test between Negro and white groups.

The correlations for all grades between the LA test and the other variables in the study are shown in Table 16. These correlations are miniscule and indicate that virtually none of the variance in the other

- - - - - - - - - - - - - - -
Insert Table 16 about here
- - - - - - - - - - - - - - -

tests is attributable to Ss' ability to listen, attend, and follow instructions. This is especially true for the Memory for Numbers test, which resembles the LA in the skills it demands, except, of course, for the memory aspect of the former. No further use need be made of the LA test, since covariance adjustment of group means or correlations among other tests would be so small as not to be detectable within the number of significant digits in these scores.

Making X's Test. Table 17 gives the statistics on this test.

- - - - - - - - - - - - - - -
Insert Table 17 about here
- - - - - - - - - - - - - - -

This is the one test in the battery on which Negro pupils obtain higher scores than white pupils at every grade level. The mean differences are statistically significant, both for absolute level of performance on Parts I and II and on the gain scores (II - I). The race difference between medians is not so striking but is in the same direction. These results show quite clearly that equally good cooperation and effort were obtained in the test situation for both white and Negro children. The lower quartile score ($Q_1$) should be a most sensitive indicator of children who are not putting out much effort, and we see that at every grade the Negro Ss equal or exceed the white Ss in performance. Covariance adjustment of means on other tests, controlling for Making X's ability, would, if anything, increase the magnitude of the white-Negro differences on the other tests. These results contradict the popular notion that Negro children have a slower "personal tempo" or are more lackadaisical in a test situation, or that their lower average performance on cognitive tasks reflects mainly a speed factor. Given a test that involves only speed but no appreciable cognitive factor, the Negro children perform as well as or better than the white children.

Memory for Numbers Test. Table 18 shows the statistics on this test. The significant white-Negro difference on the I,R, and D parts

- - - - - - - - - - - - - - -
Insert Table 18 about here
- - - - - - - - - - - - - - -

of this test are quite substantial. They can be most readily assessed

Table 16

Correlations between Listening-Attention Test and Other Variables

| Variable | White (N=1489) | Negro (N=1123) |
|---|---|---|
| Age (Mos.) | .066 | .047 |
| LT Verbal | .022 | .060 |
| LT Non V | .070 | .074 |
| Mem. I | -.001 | .054 |
| Mem. R | -.010 | .000 |
| Mem. D | .007 | .065 |
| Mem. Total | -.001 | .047 |

Table 17

Statistics on Speed and Persistence Test (Making X's)

in White and Negro Groups

Grade 4

| | White (N=542) | | | Negro (N=432) | | |
|---|---|---|---|---|---|---|
| Statistic | I | II | Gain | I | II | Gain |
| Mean | 64.46 | 77.86 | 13.40 | 74.34 | 88.66 | 14.31 |
| SD | 26.7 | 24.3 | 15.3 | 26.6 | 21.9 | 17.6 |
| $SE_M$ | 1.15 | 1.04 | 0.66 | 1.28 | 1.06 | .85 |
| Min. | 11.0 | 0.0 | -113.0 | 0.0 | 0.0 | -45.0 |
| Max. | 132.0 | 135.0 | 79.0 | 144.0 | 150.0 | 91.0 |
| Range | 121.0 | 135.0 | 192.0 | 144.0 | 150.0 | 136.0 |
| Median | 60.0 | 80.0 | 10.0 | 78.0 | 91.0 | 10.0 |
| $Q_1$ | 42.0 | 56.5 | 4.0 | 52.0 | 75.0 | 3.0 |
| $Q_3$ | 86.0 | 97.0 | 20.0 | 94.0 | 105.0 | 22.0 |

Table 17

(Continued)

Statistics on Speed and Persistence Test (Making X's)

in White and Negro Groups

Grade 5

| | White (N=498) | | | Negro (N=419) | | |
|---|---|---|---|---|---|---|
| Statistic | I | II | Gain | I | II | Gain |
| Mean | 82.44 | 94.72 | 12.28 | 82.42 | 97.47 | 15.05 |
| SD | 26.2 | 24.9 | 13.8 | 28.6 | 23.3 | 18.2 |
| $SE_M$ | 1.18 | 1.12 | 0.62 | 1.40 | 1.14 | 0.89 |
| Min. | 21.0 | 25.0 | -26.0 | 17.0 | 3.0 | -46.0 |
| Max. | 146.0 | 150.0 | 66.0 | 150.0 | 150.0 | 76.0 |
| Range | 125.0 | 125.0 | 92.0 | 133.0 | 147.0 | 122.0 |
| Median | 87.0 | 97.0 | 10.0 | 86.0 | 101.0 | 12.0 |
| $Q_1$ | 61.5 | 82.5 | 3.0 | 62.0 | 85.0 | 4.0 |
| $Q_3$ | 101.0 | 111.0 | 18.0 | 103.0 | 114.0 | 22.75 |

Table 17

(Continued)

Statistics on Speed and Persistence Test (Making X's)

in White and Negro Groups

Grade 6

| | White (N=548) | | | Negro (N=391) | | |
|---|---|---|---|---|---|---|
| Statistic | I | II | Gain | I | II | Gain |
| Mean | 95.07 | 107.27 | 12.20 | 93.37 | 108.75 | 15.38 |
| SD | 25.2 | 22.6 | 17.3 | 29.7 | 25.6 | 20.3 |
| $SE_M$ | 1.08 | 0.97 | 0.74 | 1.50 | 1.29 | 1.03 |
| Min. | 25.0 | 36.0 | -36.0 | 0.0 | 0.0 | -147.0 |
| Max. | 150.0 | 150.0 | 82.0 | 150.0 | 150.0 | 87.0 |
| Range | 125.0 | 114.0 | 118.0 | 150.0 | 150.0 | 234.0 |
| Median | 99.0 | 111.0 | 8.0 | 99.0 | 111.0 | 13.0 |
| $Q_1$ | 79.0 | 98.0 | 1.0 | 77.0 | 97.0 | 5.0 |
| $Q_3$ | 113.0 | 122.0 | 18.0 | 114.75 | 125.0 | 24.75 |

Table 18

Statistics on the Memory for Numbers Test

for Immediate (I), Repeated Series (R),

and Delayed Recall (D) in White and Negro Groups

Grade 4

| | White (N=504) | | | Negro (N=411) | | |
|---|---|---|---|---|---|---|
| Statistic | I | R | D | I | R | D |
| Mean | 21.1 | 24.7 | 22.4 | 17.2 | 21.8 | 18.4 |
| SD | 6.3 | 5.9 | 5.9 | 6.2 | 6.1 | 6.8 |
| $SE_M$ | 0.28 | 0.26 | 0.27 | 0.31 | 0.30 | 0.33 |
| Min. | 6.0 | 6.0 | 0.0 | 4.0 | 0.0 | 0.0 |
| Max. | 39.0 | 39.0 | 37.0 | 39.0 | 39.0 | 36.0 |
| Range | 33.0 | 39.0 | 37.0 | 35.0 | 39.0 | 36.0 |
| Median | 20.0 | 25.0 | 23.0 | 17.0 | 21.0 | 18.0 |
| $Q_1$ | 17.0 | 21.0 | 19.0 | 13.0 | 18.0 | 14.0 |
| $Q_3$ | 25.0 | 29.0 | 26.0 | 21.0 | 26.0 | 24.0 |

Table 18

(Continued)

Statistics on the Memory for Numbers Test

for Immediate (I), Repeated Series (R),

and Delayed Recall (D) in White and Negro Groups

Grade 5

| | White (N=477) | | | Negro (N=416) | | |
|---|---|---|---|---|---|---|
| Statistic | I | R | D | I | R | D |
| Mean | 23.5 | 26.9 | 24.4 | 18.8 | 23.0 | 19.8 |
| SD | 6.4 | 5.8 | 5.4 | 7.2 | 7.3 | 7.4 |
| $SE_M$ | 0.29 | 0.27 | 0.25 | 0.35 | 0.36 | 0.36 |
| Min. | 8.0 | 7.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Max. | 39.0 | 39.0 | 39.0 | 39.0 | 38.0 | 38.0 |
| Range | 31.0 | 32.0 | 39.0 | 39.0 | 38.0 | 38.0 |
| Median | 23.0 | 27.0 | 25.0 | 18.0 | 23.0 | 20.0 |
| $Q_1$ | 19.0 | 23.0 | 21.0 | 14.0 | 19.0 | 15.0 |
| $Q_3$ | 27.0 | 31.0 | 28.0 | 23.0 | 27.0 | 25.0 |

Table 18

(Continued)

Statistics on the Memory for Numbers Test
for Immediate (I), Repeated Series (R),
and Delayed Recall (D) in White and Negro Groups

Grade 6

| Statistic | White (N=511) I | R | D | Negro (N=388) I | R | D |
|---|---|---|---|---|---|---|
| Mean | 24.8 | 28.3 | 25.5 | 19.97 | 24.9 | 22.2 |
| SD | 6.0 | 5.4 | 5.4 | 6.8 | 6.7 | 6.6 |
| $SE_M$ | 0.27 | 0.24 | 0.24 | 0.34 | 0.34 | 0.34 |
| Min. | 5.0 | 13.0 | 0.0 | 1.0 | 6.0 | 0.0 |
| Max. | 39.0 | 39.0 | 39.0 | 39.0 | 39.0 | 39.0 |
| Range | 34.0 | 26.0 | 39.0 | 38.0 | 33.0 | 39.0 |
| Median | 24.0 | 28.0 | 26.0 | 19.0 | 24.0 | 23.0 |
| $Q_1$ | 21.0 | 25.0 | 22.0 | 15.0 | 21.0 | 18.0 |
| $Q_3$ | 29.0 | 32.0 | 29.0 | 24.0 | 29.0 | 27.0 |

by conversion to sigma units based on the SD in the white group, as shown in Table 19. The overall difference is .67σ in the population.

---

Insert Table 19 about here

---

The group difference is consistently less for the Repeated Series condition, but does not differ consistently for the others.

<u>Lorge-Thorndike Intelligence Test</u>. Table 20 shows the statistics on Lorge-Thorndike IQs. Table 21 shows the mean white-Negro difference expressed in sigma units based on the white SD. The small disparity

---

Insert Tables 20 and 21 about here

---

between means and medians in Table 20 for both white and Negro groups indicates that their respective IQ distributions do not depart appreciably from normality. The high scores of the white group in this school population are responsible for the large sigma difference between the racial groups. In the general population the white and Negro mean IQs differ by approximately 1σ.

<u>Comparison of Racial Group Means on IQ and Memory Span</u>. Tables 19 and 21 provide the basis for comparing the white and Negro groups on memory span and intelligence. The Intelligence/Memory ratio of the sigma differences for grades 4, 5, and 6 are 2.56, 2.03, and 2.76 for Verbal IQ and 2.63, 2.35, and 2.71 for Nonverbal IQ. The combined grade ratios of the sigma differences for IQ/Memory are 2.43 for Verbal and 2.54 for Nonverbal. Overall, the white-Negro IQ difference is 2.5 times greater than the white-Negro difference in total Memory score; or conversely, the white-Negro difference on memory ability is only 40% as great as the difference in IQ.

<u>Correlations Between IQ and Memory Test</u>

Because some children were not present on every one of the days on which tests were administered, the correlations among tests are based on slightly less than the complete sample summarized in the preceding tables. Table 22 summarizes the Memory and Intelligence raw scores for the groups used in the correlational analysis

---

Insert Table 22 about here

---

and shows the white-Negro mean differences in sigma units (based on the white SD). Both groups are within 1 month of 11 years of age. The Verbal IQs corresponding to the raw score means for whites and Negroes are 113 and 91; the Nonverbal IQs are 113 and 92, respectively.

Table 23 shows the correlations (Pearson *r*) among the IQ and Memory variables. Also shown are the significance levels for the

---

Insert Table 23 about here

---

Table 19

Average White-Negro Differences in Sigma Units

(Based on White SD) on Memory for Numbers Test

| Subtest | Grade | | | Mean |
|---|---|---|---|---|
| | 4 | 5 | 6 | |
| Immediate Recall | .62 | .73 | .80 | .72 |
| Repeated Series | .49 | .67 | .63 | .60 |
| Delayed Recall | .67 | .85 | .61 | .71 |
| Mean | .59 | .75 | .68 | .67 |

Table 20

Statistics on Lorge-Thorndike Verbal and Nonverbal IQs in White (W) and Negro (N) Groups

| Statistic | Grade 4 | | | | Grade 5 | | | | Grade 6 | | | | Combined Grades | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Verbal | | Nonverbal | | Verbal | | Nonverbal | | Verbal | | Nonverbal | | Verbal | | Nonverbal | |
| | W | N | W | N | W | N | W | N | W | N | W | N | W | N | W | N |
| N | 492 | 373 | 507 | 412 | 469 | 391 | 472 | 389 | 528 | 359 | 544 | 403 | 1489 | 1123 | 1523 | 1204 |
| Mean | 118.62 | 94.35 | 119.96 | 95.27 | 117.43 | 92.34 | 118.90 | 93.70 | 119.06 | 91.82 | 121.65 | 97.10 | 118.40 | 92.84 | 120.24 | 95.37 |
| SD | 16.08 | 13.98 | 15.91 | 15.49 | 16.54 | 14.42 | 14.35 | 15.49 | 14.46 | 13.17 | 13.35 | 15.23 | 15.68 | 13.91 | 14.59 | 15.46 |
| $SE_M$ | .725 | .727 | .706 | .763 | .764 | .729 | .661 | .785 | .629 | .695 | .573 | .759 | .406 | .415 | .374 | .445 |
| Max. | 150 | 145 | 150 | 143 | 150 | 148 | 149 | 138 | 150 | 140 | 145 | 143 | 150 | 148 | 150 | 143 |
| Min. | 68 | 54 | 58 | 57 | 69 | 53 | 71 | 52 | 66 | 57 | 68 | 50 | 66 | 53 | 58 | 50 |
| Range | 82 | 91 | 92 | 86 | 81 | 95 | 78 | 86 | 84 | 83 | 77 | 93 | 84 | 95 | 92 | 93 |
| Median | 120 | 94 | 122 | 95 | 119 | 93 | 119 | 93 | 121 | 91 | 124 | 97 | 119 | 92 | 120 | 94 |

Table 21

Average White-Negro Differences in Sigma Units (based on White SD) on Lorge-Thorndike Intelligence Test, Verbal and Nonverbal

| Grade | Verbal | Nonverbal |
|---|---|---|
| 4 | 1.51 | 1.55 |
| 5 | 1.52 | 1.76 |
| 6 | 1.68 | 1.84 |
| Combined | 1.63 | 1.70 |

Table 22

Raw Score Means and SDs on Intelligence and Memory Tests

and Mean White-Negro Differences in Sigma Units

for Groups Used in Correlations

| Test | White (N=1489) | | Negro (N=1123) | | $(\overline{W}-\overline{N})/\sigma_W$ |
|---|---|---|---|---|---|
| | M | SD | M | SD | |
| Age (Mos.) | 131.23 | 10.89 | 132.61 | 11.24 | -.13 |
| Intelligence | | | | | |
| Verbal | 69.85 | 12.56 | 46.24 | 16.88 | 1.88 |
| Nonverbal | 63.12 | 10.83 | 43.47 | 14.50 | 1.81 |
| Memory | | | | | |
| Immediate | 23.33 | 6.41 | 18.75 | 6.61 | .71 |
| Repeat | 26.89 | 5.81 | 23.40 | 6.56 | .60 |
| Delay | 24.25 | 5.76 | 20.29 | 6.73 | .69 |
| Total | 74.48 | 15.58 | 62.45 | 16.82 | .77 |

Table 23

Correlation Coefficients (Decimals Omitted) among Intelligence and Memory Tests (Negroes above diagonal, Whites below)

| Tests | 1<br>V | 2<br>NV | 3<br>I | 4<br>R | 5<br>D | 6<br>Tot. |
|---|---|---|---|---|---|---|
| 1. Verbal IQ | | 728 | 362 | 362 | 351 | 420 |
| 2. Nonverbal IQ | 739 | | 306 | 323 | 316 | 372 |
| 3. Memory - I | 395 | 376 | | 592 | 563 | 841 |
| 4. Memory - R | 416 | 406 | 651 | | 617 | 862 |
| 5. Memory - D | 398 | 371 | 590 | 626 | | 855 |
| 6. Memory - Tot. | 466 | 443 | 874 | 873 | 847 | |

White N = 1489

Negro N = 1123

Significance of Differences ($r_W - r_N$)

Exact 1-tailed P values)

| | | Intelligence V | Intelligence NV |
|---|---|---|---|
| Memory | I | .15 | .02 |
| | R | .05 | .01 |
| | D | .07 | .06 |
| | Tot. | .07 | .02 |

differences between the white and Negro correlations. A one-tailed test is appropriate since the theory predicts higher correlations between IQ and digit memory for the white (or higher SES) than for the Negro (or lower SES) group. All the differences are in the predicted direction. It is interesting that the largest differences are found for the Lorge-Thorndike Nonverbal intelligence scores, probably because it is a more pure measure of Level II ability than the more culturally loaded Verbal test.

The correlations in Table 23 cannot, however, be properly interpreted with reference to the hypothesis under consideration without taking into account group differences in variance on the intelligence and memory measures. We must ask, Do the correlations differ in the white and Negro samples because of group differences in variability? To answer this, the correlations must be corrected for restriction of range, which in effect equalizes the variances of the two groups. The method is explicated by Guilford (1956, pp. 320-321). In this case the correction was applied to the correlations in the white group. The crucial correlation with respect to our hypothesis is that between intelligence (Level II) and memory (Level I), so we should look at the correlations between total memory score and the Verbal and Nonverbal intelligence scores. The corrected $\underline{r}$ between Total Memory and Lorge-Thorndike Verbal is .610 for whites vs. .420 for Negroes. The difference is highly significant ($\underline{z}$ = 6.59, while for a one-tailed test a $\underline{z}$ of only 3.61 is required for significance at the .0001 level). The corrected $\underline{r}$ between Total Memory and Lorge-Thorndike Nonverbal is .585 for the whites vs. .372 for the Negroes, also a highly significant difference ($\underline{z}$ = 7.07).

Does the reliability of scores affect the differences between white and Negro correlations? We can correct for attenuation for the Memory total score. The best reliability estimate of the total score in these samples is the average correlation among the three subtests (I, R, D), boosted by the Spearman-Brown formula for a test three times as long. The resulting reliability estimates are .83 for whites and .81 for Negroes. Using these to correct for attenuation, the correlations between Total Memory and Lorge-Thorndike Verbal become .67 for whites vs. .47 for Negroes ($\underline{z}$ = 7.73), and the $\underline{r}$s between Memory and Lorge-Thorndike Nonverbal are .64 for whites vs. .41 for Negroes ($\underline{z}$ = 8.11). Thus correction for attenuation accentuates the difference. The correction for attenuation did not include Lorge-Thorndike reliability, which is close to .90 in the normative population at these grade levels. There is no reason to believe there would be a significant difference in reliability for Negro and white pupils, and the fact that the Verbal and Nonverbal tests intercorrelate .74 and .73 for whites and Negroes, respectively, makes it reasonable to assume that the reliabilities do not differ in the two groups.

We can examine the effects of age on these correlations by partialing out age in months. The correlations of the key variables with chronological age in months are as follows:

| | White | Negro |
|---|---|---|
| Verbal | .216 | .174 |
| Nonverbal | .231 | .223 |
| Total Memory | .147 | .124 |

With age partialed out of the correlations between Total Memory and Intelligence scores, the partial correlations for the Verbal test are .66 for whites vs. .45 for Negroes ($z = 7.66$), and for the Nonverbal test .63 for whites vs. .40 for Negroes ($z = 8.14$). Thus, although partialing out age lowers all the correlations slightly, it does not change the overall picture appreciably or alter the conclusions or the level of significance on which they are based.

The hypothesis that Level I and Level II tests are more highly correlated in the middle SES than in the lower SES population (in this study white vs. Negro) is thus confirmed at a high level of significance. According to our theory, the differences should be even larger for socioeconomically more extreme groups. In this study both the white and Negro groups, while representing a mean SES difference, contain a large range of SES levels with considerable overlap between the groups, so that, if anything, the results are attenuated with respect to the hypothesis. Subsequent studies will investigate the hypothesis with respect to SES levels <u>within</u> racial groups.

<u>Regression of Memory on Intelligence</u>. Probably the most informative way of looking at the relationship between the Level I (Memory) and Level II (intelligence) tests is in terms of the regression of the one variable on the other. First, let us look at the regression lines for both SES groups. The main features of the model, as shown in Figure 8, are (1) the difference between the SES means on the

Insert Figure 8 about here

Level II test ($\bar{X}_L$ vs. $\bar{X}_M$); (2) the lack of difference between the SES means on the Level I test ($\bar{X}_{L M}$); and (3) the difference in the angles between the Level I and Level II regression lines (the angles for the lower and middle class are designated l and m). (The cosine of this angle is the correlation between Level I and Level II.) Given these hypothetical conditions, and assuming linearity of regression, these are the regression lines that would result. In order to simplify Figure 8, we can remove the lines showing the regression of Level II on Level I. The result is Figure 9, showing only the regression of Level I on Level II. It can be seen that this looks very much like

Insert Figure 9 about here

Figure 5, earlier in this report, which should not be surprising, since the theory was formulated to comprehend the empirical phenomena

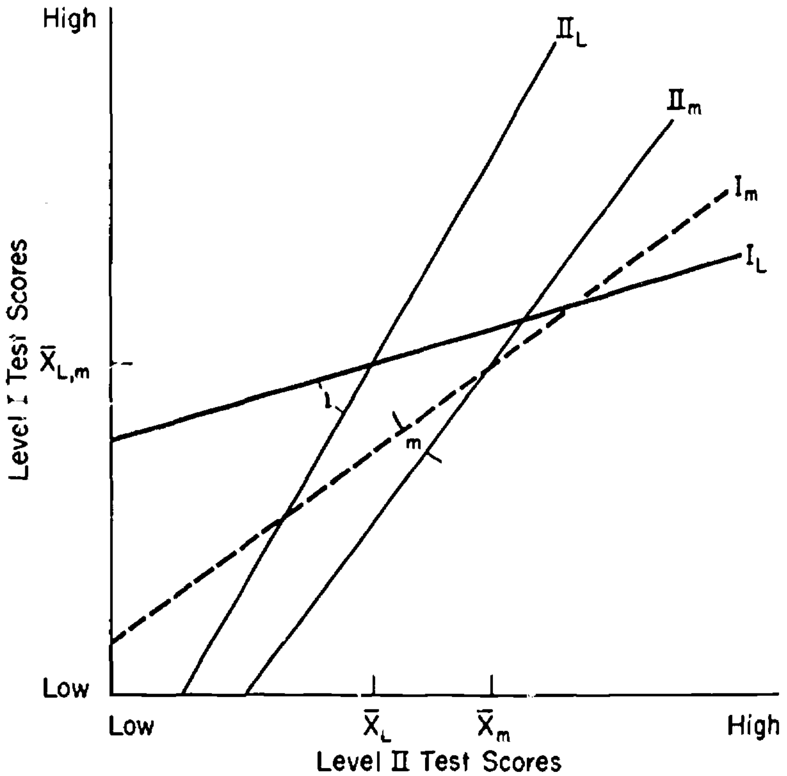



Figure 8. Hypothetical regression lines for relationship between Level I and Level II abilities in middle class (M) and lower class (L) populations.

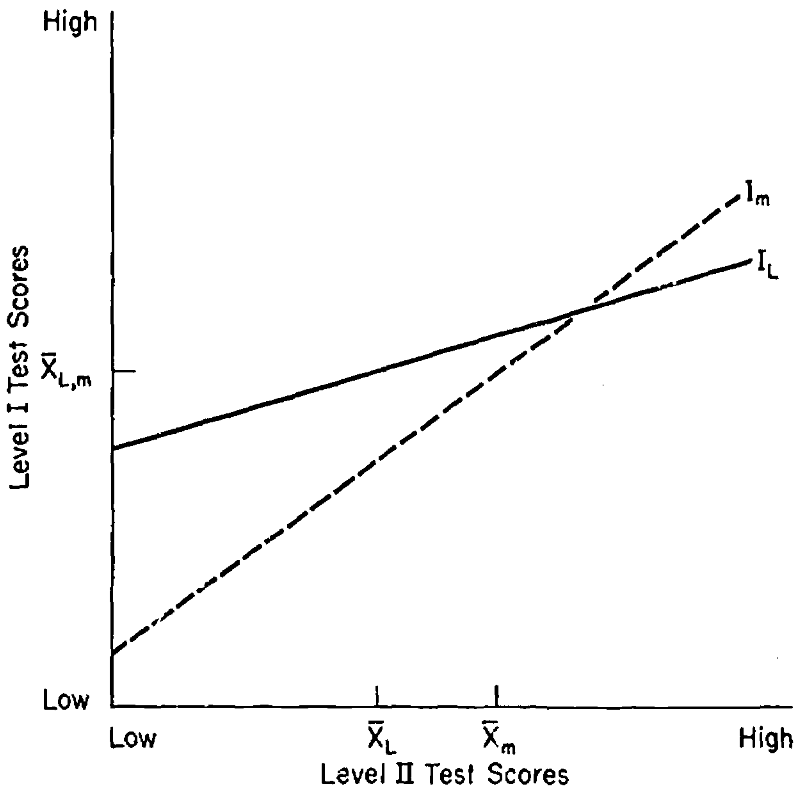



Figure 9. Hypothetical regression of Level I ability on Level II ability in middle and lower class populations.

summarized in Figure 5. But the data represented in Figure 5 were based on groups selected for being high or low on IQ (60-80 vs. 100 and above) and high or low on SES. In the present study we can now observe the actual regression lines based on an entire school population in grades 4 through 6. These regression lines, based on raw scores for both the memory and intelligence tests, are shown in Figures 10 and 11 for the Lorge-Thorndike Verbal and Nonverbal scores, respectively. Tests of the linearity of regression show no significant

- - - - - - - - - - - - - - - - - -

Insert Figures 10 and 11 about here

- - - - - - - - - - - - - - - - - -

departure from linearity throughout the entire range of scores in both white and Negro groups. The length of the regression lines corresponds to the full range of scores of pupils in regular classes. (Children in special classes were not included in this study.)

The picture is essentially the same for both the Verbal and Nonverbal tests. The regression lines for whites and Negroes cross at a point equivalent to a Lorge-Thorndike IQ of 98 on both V and NV tests. That is to say, at IQ 98, white and Negro children on the average have exactly the same memory scores. As the IQ goes <u>below</u> 98, Negro children increasingly excel white children in memory score, on the average; and as the IQ goes above 98, white children increasingly excel Negro children in memory performance. This would mean that, on the average, the white child below approximately IQ 98 has a poorer memory span than his Negro counterpart in IQ, and that the difference increases, in favor of the Negro child, the lower the IQ. In terms of nationwide IQ norms the approximately 80 to 85 percent of Negro children who fall in this range excel the 50 percent of white children in this range. The results in Figures 10 and 11, however, are at variance with the model as shown in Figures 8 and 9 in the fact that the two SES groups (Negro vs. white) differ in mean digit span, even when the digit span scores are read off the regression lines for IQs 85 and 100, which are approximately the Negro and white mean IQs on a nationwide basis.

How much overall intellectual advantage or disadvantage is associated with a memory span higher or lower than the IQ is not known. It may well be a greater <u>advantage</u> to have a higher memory span than the IQ when the IQ is low than it is a <u>disadvantage</u> to have a lower memory span than the IQ when the latter is high. The answer will have to await subsequent studies which will examine the multiple regression of performance in various scholastic subjects on digit memory and intelligence test scores.

Figures 10 and 11 make it clear that in comparing lower and higher SES groups, their respective means on the intelligence test scale will determine whether there are or are not differences between them on Level I tests and will determine the <u>direction</u> of the difference.

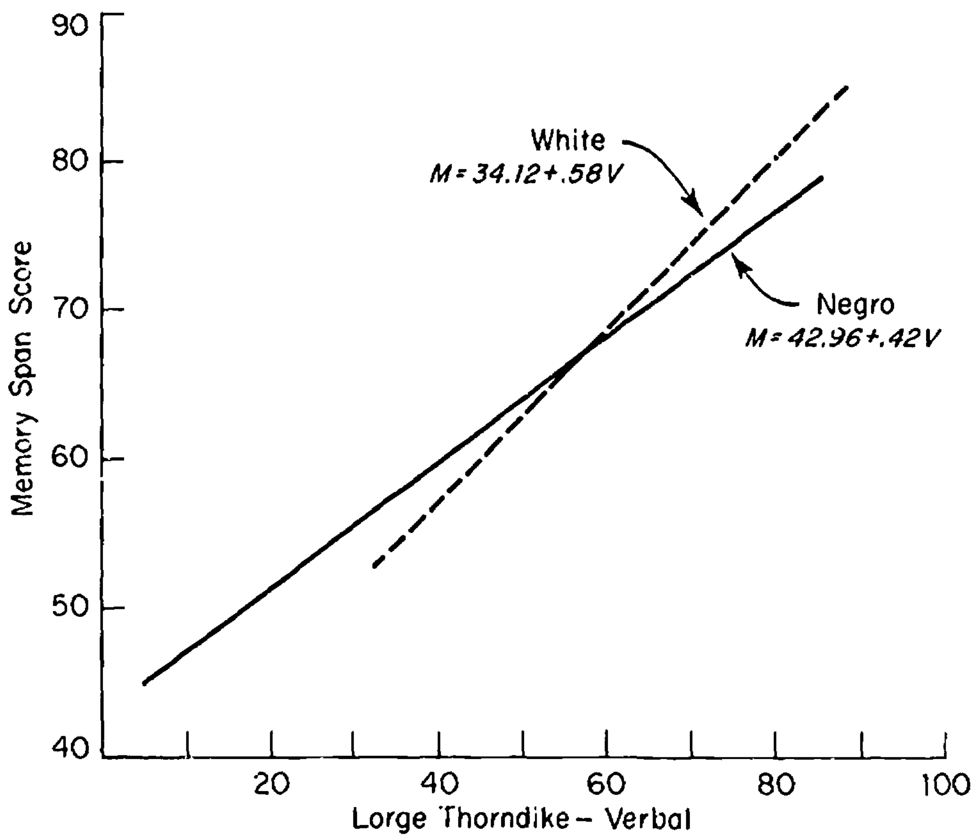



Figure 10. Regression of memory scores on Lorge-Thorndike Verbal Intelligence Scale raw scores in white and Negro children in grades 4 to 6.

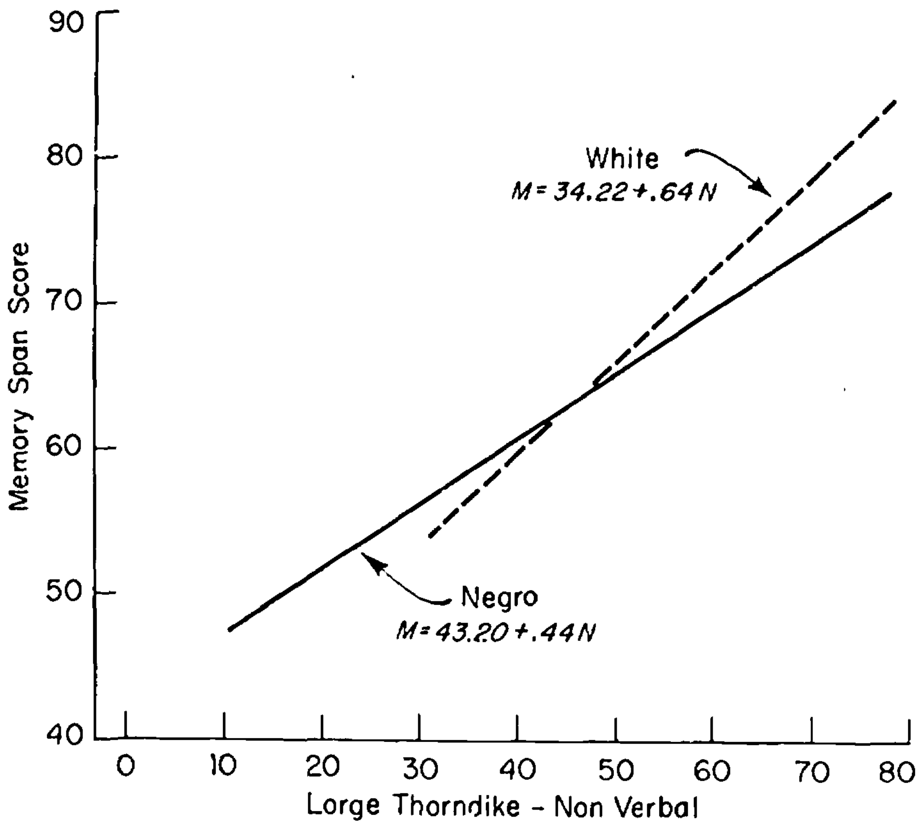



Figure 11. Regression of memory scores on Lorge-Thorndike Nonverbal Intelligence Scale raw scores in white and Negro children in grades 4 to 6.

Regression of Intelligence Test Scores on Memory. We now reverse the axes and look at the regression of Lorge-Thorndike Verbal and Nonverbal scores on the memory test, as shown in Figures 12 and 13. These regression lines present a very different picture indeed from

Insert Figures 12 and 13 about here

those in Figures 10 and 11. They correspond fairly well to the hypothetical regression lines ($II_M$ and $II_L$) shown in Figure 8, although the latter show some convergence, which is required by the theory. The regression lines in Figures 12 and 13 represent the actual raw data of the present study, in which the variances of the racial groups are unequal, which is responsible for the non-converging regression lines, despite different correlations between Level I and Level II for the white and Negro groups, as was shown in the previous section on the correlations. Since the slope of the regression line of y on x is $r_{xy}(\sigma_y/\sigma_x)$, it will be affected by inequalities in the sigmas of the white and Negro groups. If the regression lines were corrected for unequal variances the results would necessarily conform more closely to the model, since the slope of the white's regression line would be steeper relative to the Negro's. The difference, however, would not be great, and it seems preferable at this point to show the raw results without any statistical adjustments.

What the regression lines in Figures 12 and 13 show, of course, is that at any level of memory span there is a constant average white-Negro intelligence difference (both Verbal and Nonverbal) of something more than 1 SD. The white-Negro difference in memory span for any given IQ is relatively small and in favor of Negroes for IQs below 98 (Figures 10 and 11). The reverse (Figures 12 and 13) is very different: the white-Negro IQ difference is almost uniformly large at every level of memory span. Only Negroes in the highest quartile of memory span obtain Lorge-Thorndike scores as high as whites who are in the lowest quartile in memory span. In other words, in this population if white and Negro children are matched on IQ, they will be similar in memory span, but if matched on memory span they will differ, on the average, more than 1 SD in IQ. This suggests a hierarchical relationship between memory span and intelligence. That is, high intelligence indicates high memory ability to a much stronger degree than high memory ability indicates high intelligence. This is in line with the "necessary-but-not-sufficient" formulation of the relationship between Levels I and II. The theory postulates that Level I ability is necessary but not sufficient for the development of Level II ability. What this means in terms of the data is just what we see in comparing Figures 10 and 11 with Figures 12 and 13, plus one other feature of the correlation scatter diagram which is hypothesized by Figure 3 in the theoretical introduction. The hypothesis illustrated in this exaggerated figure is the prediction of a broader scatter of memory ability at lower levels of IQ than at higher levels. In other words, the scatter or dispersion around the regression line of memory on intelligence should decrease

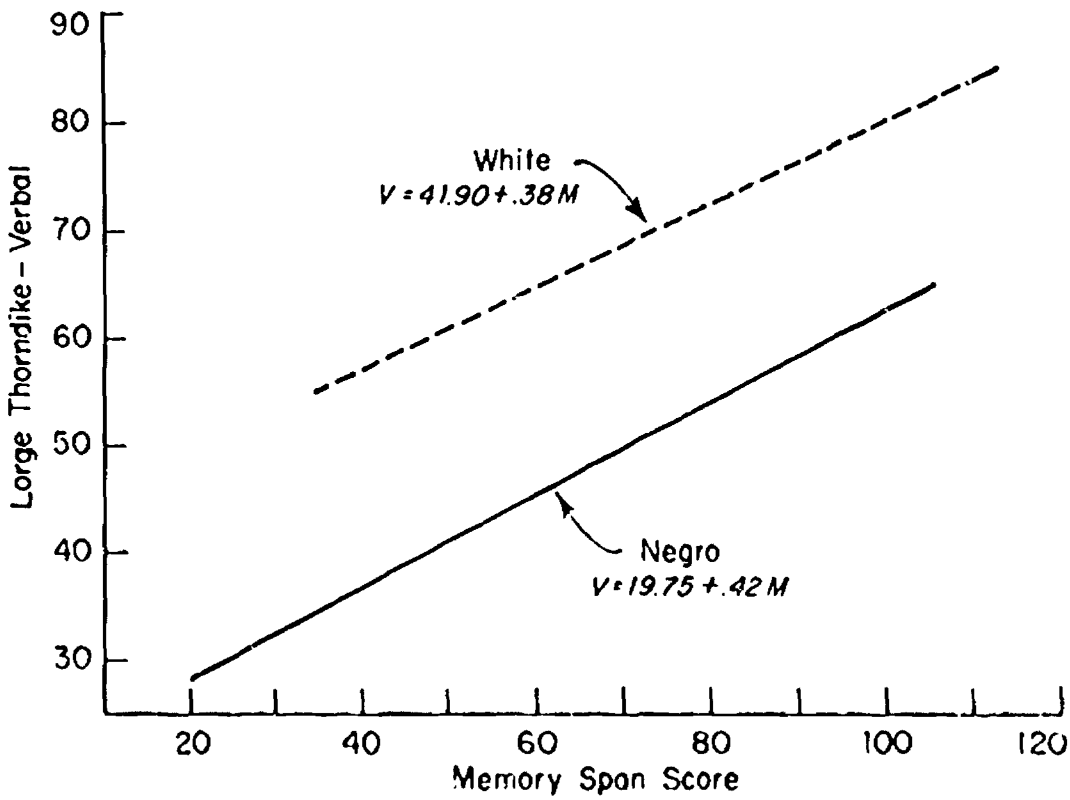



Figure 12. Regression of Lorge-Thorndike Verbal raw scores on memory scores in White and Negro children in grades 4 to 6.

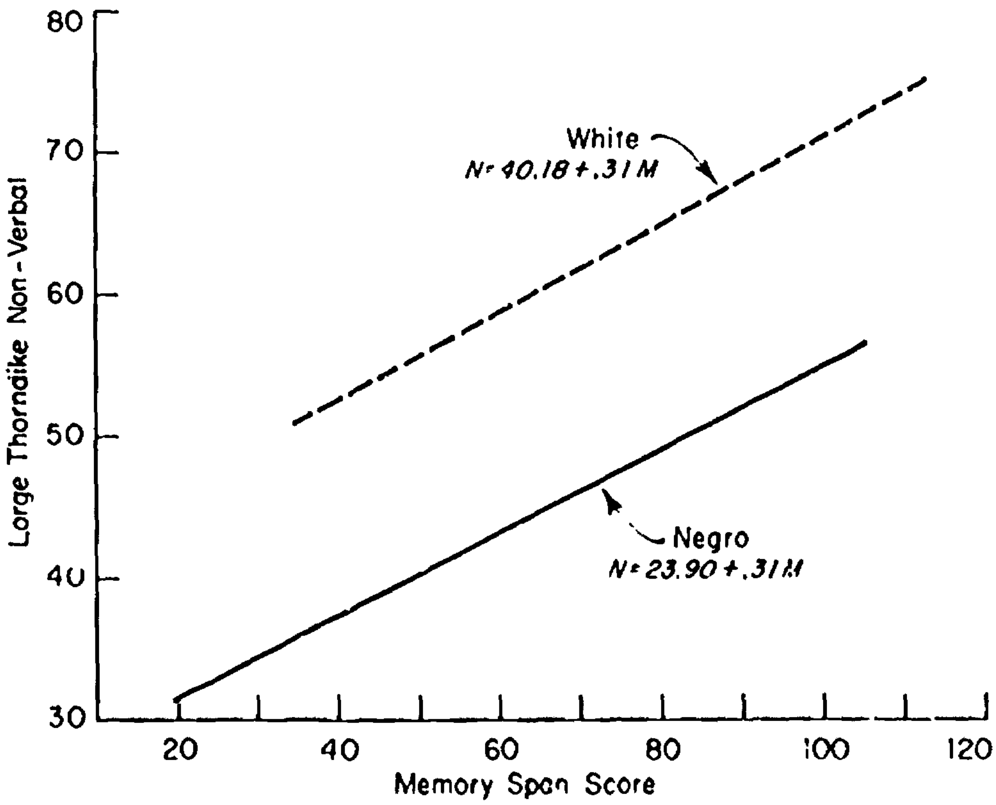



Figure 13. Regression of Lorge-Thorndike Nonverbal raw scores on memory scores in white and Negro children in grades 4 to 6.

as intelligence increases.

Dispersion of Memory Ability (Level I) as a Function of Intelligence (Level II). This hypothesis is best tested by examination of the standard error of estimate around the regression line of memory on intelligence. The standard error of estimate in this case is the standard deviation of memory scores for any given intelligence test score. The theory predicts that the standard error of estimates should be greater at the lower end of the intelligence scale than at the higher end, and more so in the low SES than in the middle SES group. (See graphic representation of this hypothesis in Figure 3.) Bartlett's test for homogeneity of variances was performed on the data and showed differences significant beyond the .01 level in the memory score variances as a function of IQ level. So differences in the standard error of estimates ($SE_E$) are significant, but the important question concerns the trend of the differences; according to the theory, the $SE_E$ should decrease with increasing IQ. The trend can be examined graphically by plotting $SE_E$ as a function of Lorge-Thorndike Verbal and Nonverbal intelligence scores, as shown in Figure 14. A smoothed line, based on a moving average of every three adjacent data points, is presented

Insert Figure 14 about here

to show the trend more clearly, since the $SE_E$ is quite erratic. The predicted downward trend in $SE_E$ is clearly apparent and more pronounced in the Negro sample, as also predicted. But it is also considerably less regular and clear-cut than was the impression gained from previous studies based on smaller and more extreme groups, and the trend is evident only on the Lorge-Thorndike Nonverbal test. At this point one can only speculate as to the reason for this difference. It is likely that the Nonverbal test is less culture-loaded and not dependent on reading ability and is therefore a more pure measure of Level II ability. Throughout these studies the nonverbal test has consistently conformed more closely to theoretical predictions for Level II than the Verbal test. The present results suggest that while the hypothesized "necessary-but-not-sufficient" relationship between Level I and Level II abilities is valid, it operates within very broad limits. On the average, however, prediction from intelligence to memory span is better than prediction from memory span to intelligence if one does not take SES or racial group into account. This was illustrated in Figures 10 through 13. Figures 10 and 11 show that, on the average, one would not be far off in predicting memory span from the intelligence test scores without taking the racial group membership of individuals into account. Figures 12 and 13, on the other hand, show that the average prediction of intelligence from a knowledge of the memory score depends strongly upon the racial (or SES) group.

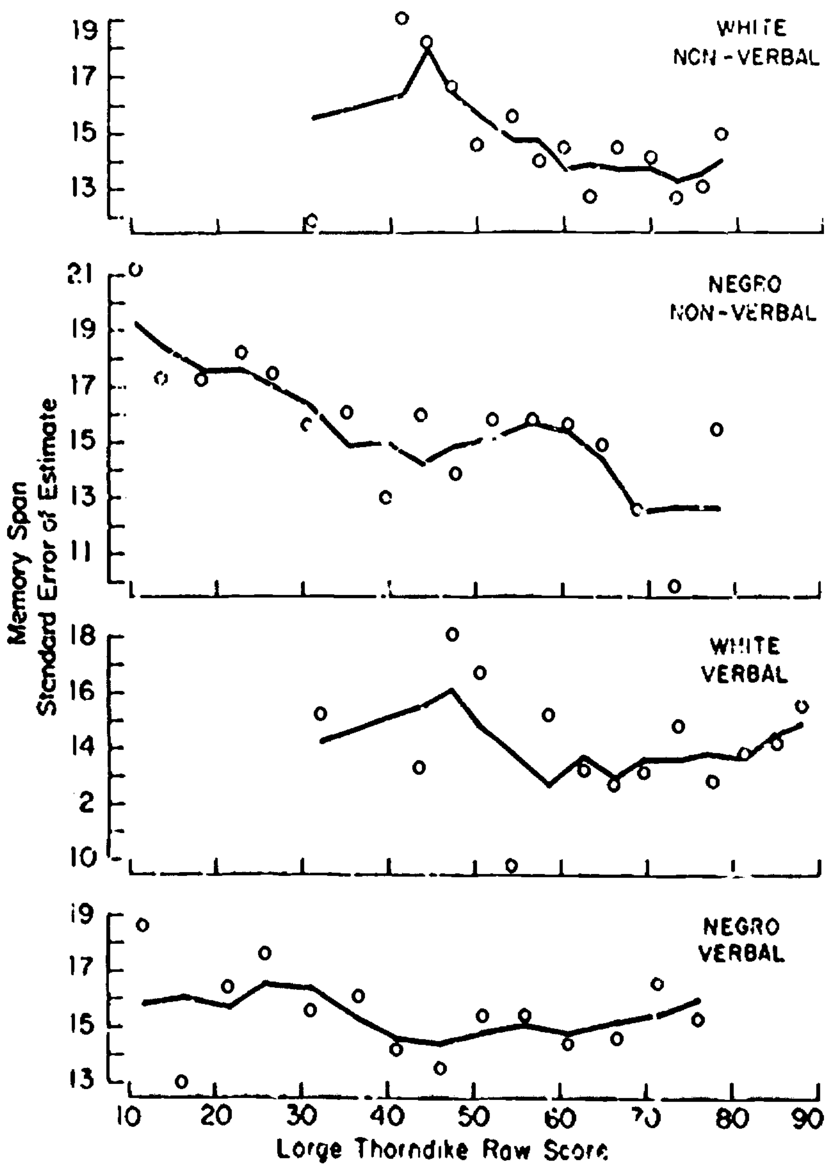



Figure 14. Memory score dispersion (standard error of estimate) as a function of Lorge-Thorndike raw scores in white and Negro groups in grades 4 to 6.

Relationship of the Draw-a-Man Test

to Level I and Level II

Arthur R. Jensen

The Harris-Goodenough Draw-a-Man Test (DMT) requires the child to draw a picture of a man, which is scored for various features on a mental maturity scale. The point scores can be converted to mental age and IQ. Since there have been claims that this test is more culture-fair and discriminates less between lower and middle class children, the present study was intended to determine whether the DMT is more highly related to Level I or Level II ability.

## Method

### Subjects

Ss were tested in intact classes from kindergarten through grade 6 in two schools: the Low SES school was in a relatively poor neighborhood and nearly all the children were Negro; the Middle SES school was in an all-white middle and upper-middle-class neighborhood. Grade 2 was omitted, since they were taking part in another study.

The DMT test was group administered by a trained psychometrist in accord with the standard instructions given in the manual. All the tests were scored blind (i.e., no identification as to race or SES was given) by a psychologist experienced in the use of the DMT.[1]

Raven's Colored Progressive Matrices and the Memory for Numbers test were administered individually to 50 children in grades 4, 5, and 6 in each school, in order to determine the correlations among the DMT, Raven, and Memory tests in both the low and middle SES groups.

## Results

Table 24 gives the means and SDs of DMT IQs at all grade levels.

- - - - - - - - - - - - - - -
Insert Table 24 about here
- - - - - - - - - - - - - - -

These results indeed show smaller differences between the SES groups at every grade level than is generally found on other tests of intelligence. For example, the difference between these schools is close to 2 SDs on the Lorge-Thorndike intelligence test. The DMT, on the other hand, shows differences which range between .44 and .88 in sigma units, that is, differences less than half as large as those found with conventional IQ tests. But the results shown in Table 24

[1] We are indebted to Dr. Gere Watkins for his scoring of these tests.

Table 24

IQs of Low SES and Middle SES Groups

on the Harris-Goodenough Draw-a-Man Test

| Grade | Middle SES N | M | SD | Low SES N | M | SD | $(\bar{X}_M-\bar{X}_L)/SD_M$* |
|---|---|---|---|---|---|---|---|
| K | 77 | 92.18 | 10.40 | 121 | 83.83 | 9.82 | 0.80 |
| 1 | 93 | 94.43 | 12.33 | 147 | 88.96 | 12.12 | 0.44 |
| 3 | 122 | 94.31 | 13.55 | 126 | 84.71 | 12.33 | 0.71 |
| 4 | 106 | 91.61 | 11.37 | 137 | 84.39 | 12.09 | 0.64 |
| 5 | 91 | 91.20 | 10.01 | 127 | 82.38 | 11.85 | 0.88 |
| 6 | 103 | 87.85 | 9.73 | 121 | 79.88 | 11.70 | 0.82 |

*All differences are significant beyond the .01 level.

are peculiar in another respect which leads us to accept these DMT scores with caution. The IQs of the Low SES are not higher than their IQs on conventional tests such as the Lorge-Thorndike, on which the average is 85 in the Low SES school. It is the middle SES group that has below average IQs on the DMT! These children average about 115 on the Lorge-Thorndike. Thus nearly all the reduction in SES IQ difference is the result of a lowering of IQ in the middle SES group. One may wonder what kind of school population would obtain at least average IQs on the DMT if this middle SES sample does not. We have no explanation for this anomaly and find no good basis for deciding to what extent it may invalidate the SES group differences shown in Table 24.

Table 25 shows the means and SDs of the 50 Ss in each SES group on the tests used in the correlational analysis. Table 26 shows

Insert Tables 25 and 26 about here

the zero-order correlations and the partial correlations among the three tests. The partial correlations show to what degree the DMT resembles the Level I (memory) and Level II (matrices) tests independently of the correlation between Levels I and II. All the correlations are unimpressive, but the partial r with the Memory test is very low, suggesting that the DMT is more a Level II than a Level I measure. The fact that the DMT has little in common with either the Matrices or the Memory tests, which are our purest measures of Levels I and II, is shown by the fact that partialing DMT out of the correlation between Matrices and Memory only slightly lowers the correlation, as can be seen by comparing the zero-order rs with the corresponding partial rs in Table 26. Since the matrices are a good measure of "g," the general factor common to most intelligence tests, its low correlation with the DMT suggests that the latter is a rather poor measure of "g," which in our theory is practically synonymous with Level II. The factorial composition of the DMT can only be discovered through factor analysis with many more tests than were used in this study. It is clear from the present evidence, however, that the DMT seems not to be a particularly good measure of either Level I or Level II abilities.

Table 25

Means and SDs of Low and Middle SES Groups Used in

Correlational Analysis of the Draw-a-Man Test

| Test | Middle SES (N=50) M | Middle SES (N=50) SD | Low SES (N=50) M | Low SES (N=50) SD | $\frac{(\bar{X}_M-\bar{X}_L)}{SD_M}$ |
|---|---|---|---|---|---|
| Draw-a-Man | 91.84 | 12.31 | 87.08 | 15.63 | 0.39 |
| Raven's Matrices | 30.37 | 4.00 | 21.68 | 5.09 | 2.17 |
| Memory for Numbers | 234.53 | 75.74 | 176.50 | 58.56 | 0.77 |

Table 26

Correlations among Draw-a-Man, Raven's Colored Progressive Matrices, and Memory for Numbers (Low SES below Diagonal, Middle SES above)

| Test | Zero Order r's 1 | Zero Order r's 2 | Zero Order r's 3 | Partial r's 1 | Partial r's 2 | Partial r's 3 |
|---|---|---|---|---|---|---|
| 1. DMT | | .39 | .34 | | .25 | .15 |
| 2. Matrices | .52 | | .58 | .42 | | .52 |
| 3. Memory | .35 | .46 | | .15 | .35 | |

## Comparison of "Culture-Loaded" and "Culture-Fair" Tests

Arthur R. Jensen

A strictly cultural or environmental hypothesis of social class differences in intelligence holds that the differences are attributable to "culture bias" or "cultural loading" of the particular intelligence tests. All but the most naive theories in this class would acknowledge that culture bias is not a 0,1 or an "either-or" property of tests or test items. There must be degrees of culture bias for various tests, such that tests (or items) could be rank ordered on this attribute. Granted this possibility, the cultural hypothesis of SES differences should predict that tests which are more culture biased should yield larger mean differences between low and middle SES groups than tests which are less culture biased. The magnitude of the SES difference in test scores cannot itself properly be used as the criterion of culture bias, since this would be to make the independent and dependent variables one and the same.

Culture bias is doubtlessly multidimensional. That is to say, tests could be ordered differently by different criteria of culture bias which are still independent of the magnitude of SES differences in test scores. For example, one could order tests in terms of the amount of reading skill they require on the part of the subject, or in terms of amount of pictorial material characteristic of middle class culture (e.g., musical instruments, zoo animals, "fancy" furniture or tableware, etc.), or in terms of the amount of scholastic content (arithmetic, remote factual information, etc.) in the tests, and so on. The rank order of tests on these various criteria may be quite far from perfectly correlated.

It was hypothesized in the theoretical introduction to this report that at least two dimensions of test attributes are required to comprehend SES differences: culture loading and complexity. These two dimensions are primarily defined by the means by which the test items increase in difficulty. Highly culture loaded tests contain items which increase in difficulty (defined as the percent of the normative population not passing the item) by increasing the rarity of the item content. That is, the more difficult items are those calling for information with lower probability of being acquired in the culture--for example, being able to identify a picture of an aardvark as compared with a picture of a dog. The only reason that "aardvark" is more difficult than "dog" is its rarity of the word in our language and the rarity of the animal in our common experience. The items do not differ in complexity or conceptual difficulty, yet their difficulty levels in terms of *p* values (proportion of the population passing) are probably close to .01 vs. .99. Those who criticize intelligence tests as being culturally biased and therefore unfair to low SES subjects almost invariably have this criterion of culture loading in mind.

But test items can also be increased in difficulty by increasing their complexity -- the number of factors (and their degree of abstractness) that must be mentally manipulated more or less simultaneously in order to arrive at the correct answer. The contents or elements of the problems may be no more abstruse or rare for the complex problems than for the simple problems. Figure 15 illustrates this two-dimensional hypothesis. Tests are seen as vectors in the

- - - - - - - - - - - - - -

Insert Figure 15 about here

- - - - - - - - - - - - - -

two-dimensional space defined by the continua of culture loading and complexity. Tests or their items increase in difficulty along these vectors. And it is hypothesized further that the magnitude of SES differences (or the correlation of SES and test scores) is directly proportional to the length of these vectors for a given test. In Figure 15 the numbers on the vectors are proportional to their lengths. We can speculate on the location of various tests in this schemata. Test #13 could be the Peabody Picture Vocabulary Test; the items barely increase in complexity from the easiest to hardest, but do increase in rarity. Test #14 would be almost the opposite, like Raven's Progressive Matrices, which becomes progressively more difficult by increasing the complexity of the mental operations needed to arrive at the correct solution, even though the problems are made up of quite simple basic geometric forms at all levels of difficulty. Test #16 is highly loaded on both factors; such a test is the Terman Concept Mastery Test, which involves both a knowledge of scholastic-type information and the ability to figure out complex verbal analogies, similarities and differences, and the like. Tests #8 and #11 may be like the Verbal and Nonverbal parts of the Lorge-Thorndike Intelligence Tests. Tests #1 and #3 would be like forward and backward digit span. These tests can be made very difficult, but not by virtue of increasing complexity or increasing rarity of the materials. It can be seen that the complexity dimension in Figure 15 is one of increasing Level II functions, in terms of our Level I-Level II theoretical distinction. Highly complex problem solving necessarily involves Level II; it may or may not make demands on Level I ability. The reason that test #3 (backward digit span) is represented by a longer vector than test #1 (forward digit span) is that backward span involves more Level II ability, since it requires a transformation of the input. (Horn (1970) has reported that backward digit span has a higher g loading than forward span.)

The present study tests the hypothesis that intelligence tests differ along at least these two dimensions -- complexity and culture-loading -- and that various culturally disadvantaged groups may not remain in the same rank order in mean scores on tests representing different vectors in this two-dimensional space.

The study also provides a test of the culture bias hypothesis of SES differences. If low and middle SES groups are equated in performance on a culture loaded test, as by exact matching of individuals, the culture-bias hypothesis predicts that the low SES group should excel the performance of the middle SES group on a less culture loaded test.

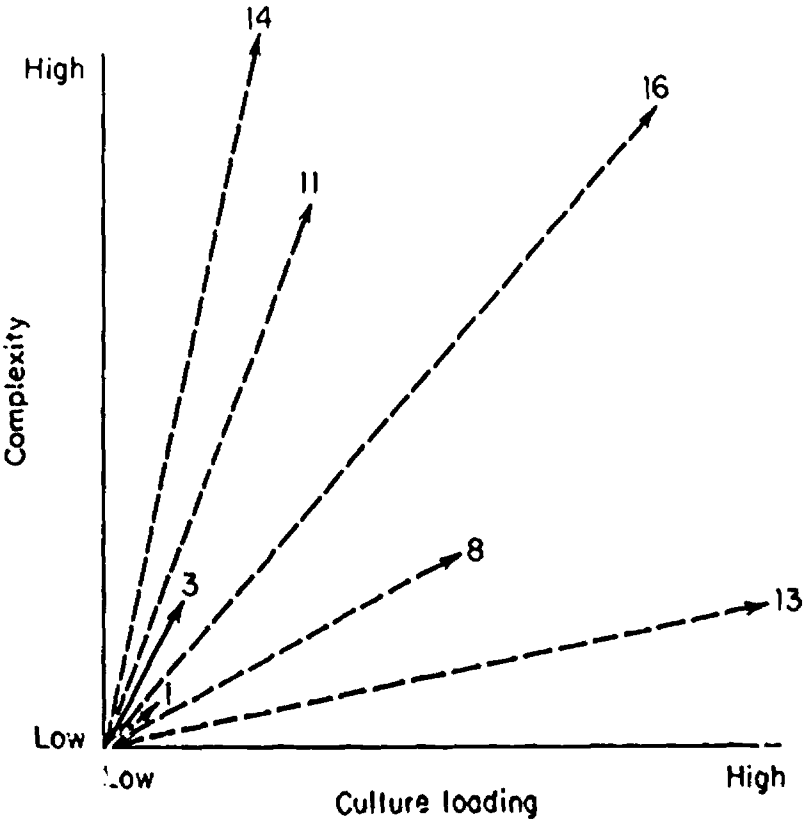



Figure 15. Hypothetical vectors proportional to social class differences for various tests in 2-dimensional space defined by complexity and culture loading of tests (or test items). The numbers are directly proportional to the lengths of the vectors.

With these hypotheses in mind, it should be instructive to compare two of the most extremely different tests with reference to Figure 5 -- the Peabody Picture Vocabulary Test (PPVT) and Raven's Colored Progressive Matrices (RCPM) in culturally disadvantaged and advantaged populations. The PPVT is an obviously culture-loaded test. The RCPM was designed to be one of the most culture-free tests of intelligence. (Of course no test stands at either end point on the culture-loading dimension.) By all commonly accepted criteria, the PPVT and RCPM stand about as far apart on the culture loading dimension as any standardized tests. Also there is little question as to the basis of the increasing difficulty in test items. The more difficult PPVT items are simply more rare; the more difficult RCPM problems, however, clearly involve more stimulus material.

## Method

### Subjects

The Ss were 1663 white, Negro, and Mexican-American children in grades kindergarten through six. The white sample (N = 638) was predominantly middle SES while the Mexican (N = 644) and Negro (N = 381) groups were predominantly lower SES. All Ss were tested individually on the Raven and the Peabody.[1] Although many of the Mexican children were bilingual and all who were tested could speak English, an English vocabulary test such as the PPVT must obviously be more culturally biased in this sample than a nonverbal test such as the Progressive Matrices.

## Results

### Rarity of Items in the PPVT

Item difficulty in the PPVT increases progressively throughout the 150 items of the test by simply increasing the rarity of the vocabulary used in connection with the pictures. To test this hypothesis it was simply necessary to plot the frequency of occurrence per million words in the English language as tabulated in the Thorndike-Lorge frequency count (Thorndike & Lorge, 1944) as a function of item difficulty. The 150 items are arranged in order of difficulty (percent not passing). The Thorndike-Lorge frequency (the G or general count) was determined for each word (in equivalent Forms A and B) and averaged over each set of 15 items. The results, as shown in Figure 16, are so absolutely clear as to need no further commentary.

Insert Figure 16 about here

[1]We are indebted to Dr. Mabel C. Purl, Director of Research and Evaluation, Riverside Unified Schools, for these data.

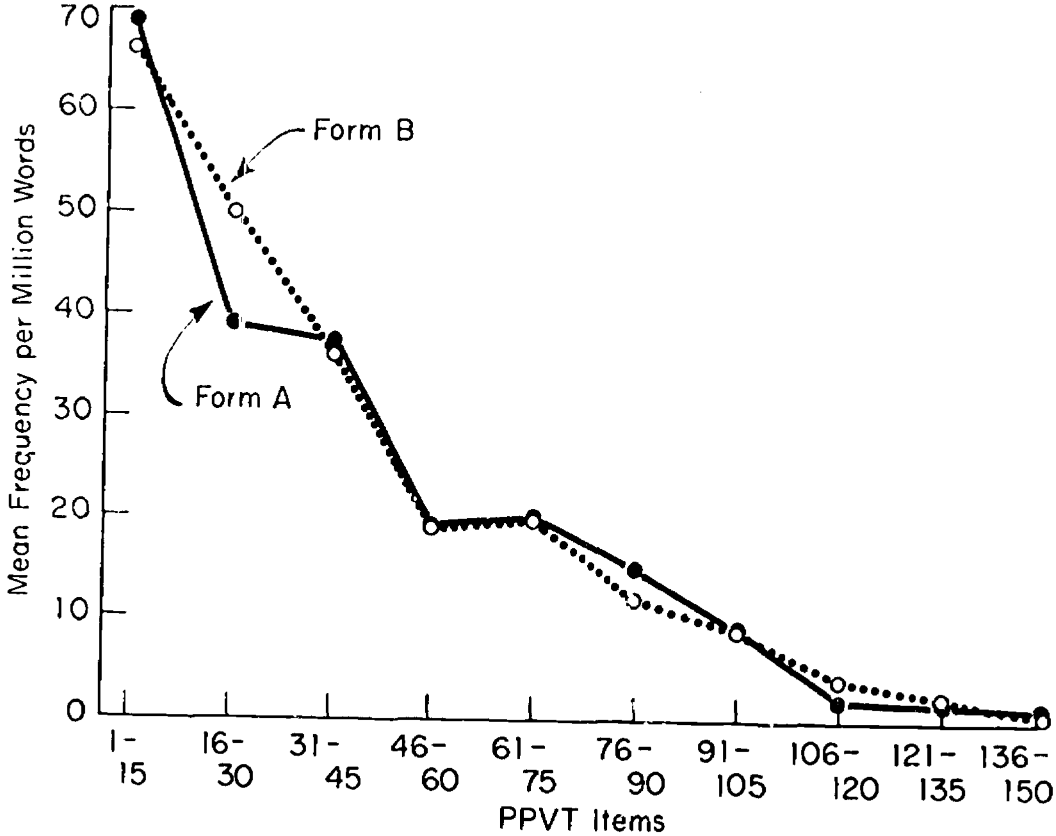



Figure 16. Mean Thorndike-Lorge word frequency of Peabody Picture Vocabulary Test items (for Forms A and B) as a function of item difficulty when items are ranked from 1 to 150 in p values based on normative population.

## Group Comparisons on Raw Scores

Figures 17 and 18 show the age trends in raw scores on the PPVT

- - - - - - - - - - - - - - - - - - -
Insert Figures 17 and 18 about here
- - - - - - - - - - - - - - - - - - -

and RCPM. The means increase with age quite linearly in the PPVT. In both tests the SDs increase only slightly. The negatively accelerated growth curve for white Ss on the RCPM is undoubtedly due to a ceiling effect imposed at higher grade levels by using only the Colored Matrices -- that is, the children's form. We have found that beyond grade 4 a small but increasing proportion of children, especially those of upper SES, attain maximum scores on the Colored Matrices. Thus the mean is slightly depressed from what it would be if the test had more "top." The trend is still quite linear throughout this age range for the other two groups, and therefore it is safe to conclude that the test slightly underestimates the group differences in intelligence beyond nine or ten years of age.

The most important feature of these two figures, however, is the fact that the relative positions of the Negro and Mexican pupils are reversed. This interaction is significant beyond the .01 level. Two tests which order the means of three groups differently must be differentiating among the groups on more than one dimension. The Mexican mean score appears to be lower primarily on the culture loading factor; the Negro score on the complexity or Level II factor. To examine this hypothesis further we must look at the intercorrelations among the tests.

## Correlations Among the Variables

The correlation between PPVT raw scores and Raven raw scores over all grades is 0.724 ($N$ = 1663). The correlation of age (in months) with PPVT raw score is .632; with Raven it is .654. The correlation between PPVT and Raven with age partialed out is 0.531. Since the reliabilities of both of these tests are close to .90, it is clear that with a correlation of only .53 they are not measuring entirely the same mental abilities. Table 27 presents the intercorrelations separately for each of the groups, and also the partial correlation between PPVT and Raven with age held constant.

- - - - - - - - - - - - - - -
Insert Table 27 about here
- - - - - - - - - - - - - - -

Using a combination both of multiple correlation (R) and partial correlation (r) tells virtually the whole story. Such an analysis is shown in Table 28. Since it seems desirable to partial out age

- - - - - - - - - - - - - - -
Insert Table 28 about here
- - - - - - - - - - - - - - -

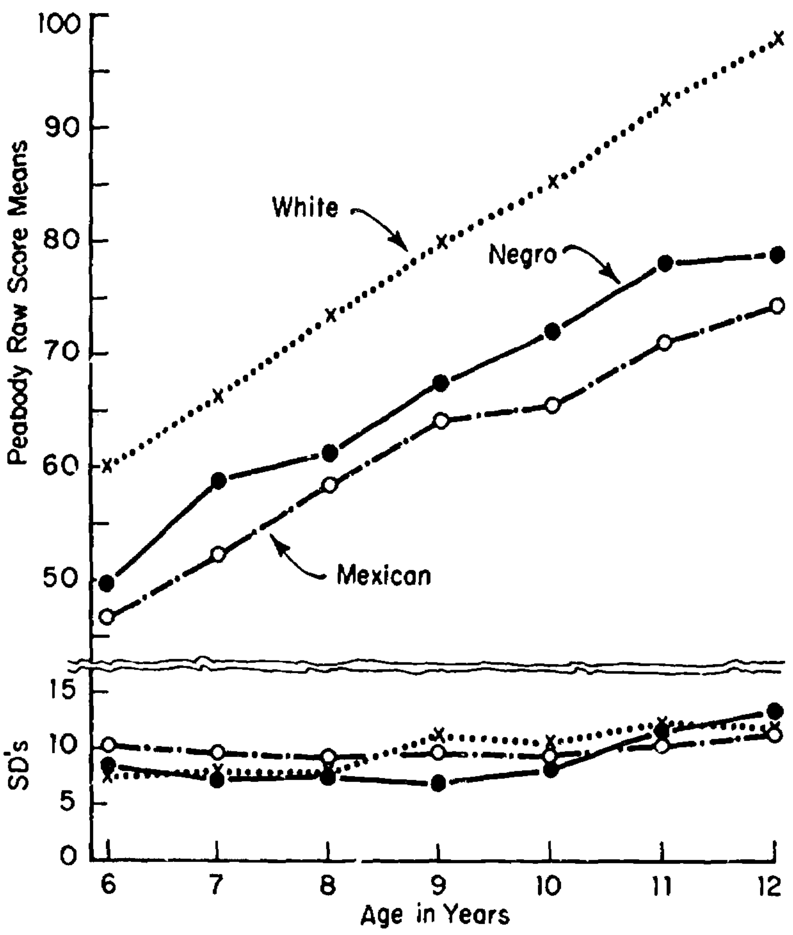



Figure 17. Peabody Picture Vocabulary Test raw scores as a function of age. Standard deviations (SDs) at each age are shown in lower part of graph.

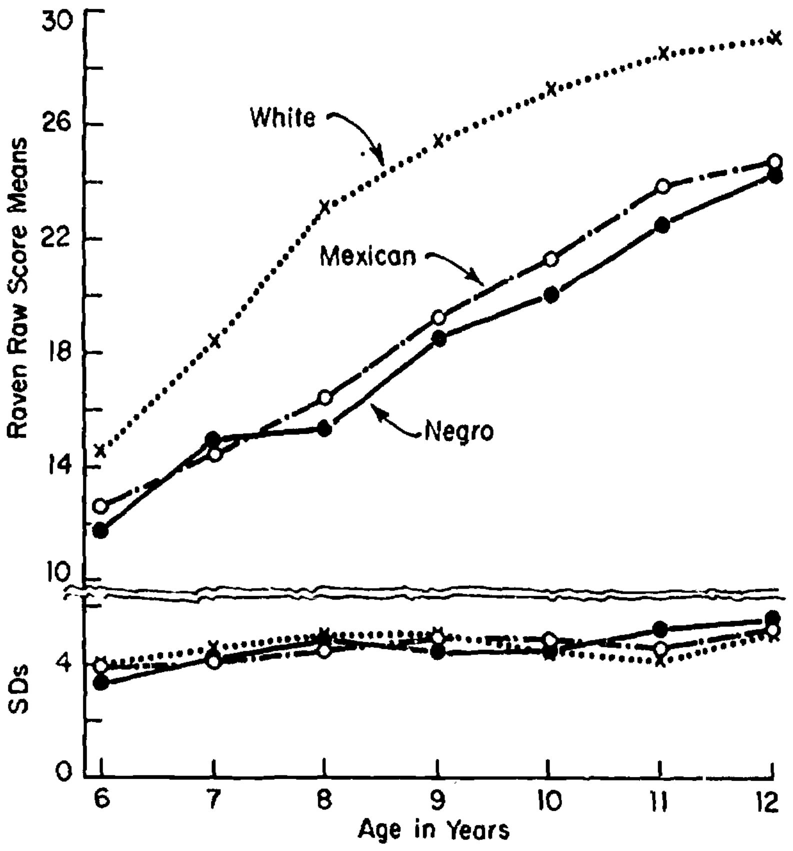



Figure 18. Raven's Colored Progressive Matrices raw scores as a function of age. Standard deviations (SDs) at each age are shown in lower part of graph.

Table 27

Correlations among Age (in months),
PPVT, and Raven's Colored Progressive Matrices
in White, Negro, and Mexican Groups

| Correlation | White (N=638) | Negro (N=381) | Mexican (N=644) |
|---|---|---|---|
| PPVT X Age | .787 | .728 | .671 |
| Raven X Age | .722 | .660 | .702 |
| PPVT X Raven | .719 | .692 | .667 |
| Partial r | | | |
| PPVT X Raven | .354 | .412 | .371 |

Table 28

Multiple and Partial Correlations between Tests and Ethnic Classification

| | Ages 5-5 to 7-6 | | | | Ages 7-7 to 9-6 | | | | Ages 9-7 to 15-0 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Criterion Groups[2] | Mult. R[1] | Partial r. PPVT | Raven | Age | Mult. R[1] | Partial r. PPVT | Raven | Age | Mult. R[1] | Partial r. PPVT | Raven | Age |
| W vs. N | .51 | .38 | .18 | -.23 | .62 | .33 | .33 | -.08 | .61 | .35 | .32 | -.18 |
| W vs. M | .63 | .55 | .04 | -.14 | .66 | .44 | .25 | -.19 | .70 | .59 | .14 | -.34 |
| N vs. M | .29 | -.28 | .13 | .00 | .24 | -.22 | .12 | .17 | .36 | -.32 | .23 | .15 |

[1] The multiple R is shrunken.

[2] In computing correlations:

White = 3,

Mexican = 2,

Negro = 1

from the correlations, and since the regression of Raven scores on age departs significantly from linearity when the total age range is considered, it was decided to divide the total sample into three groups according to age, such that within each age range, the regression of test scores on age does not depart significantly from linear regression. In this way we are able to partial out age (in months) to the maximum extent. An added advantage in analyzing the data by age groups is that it can then reveal any trends in group differences as a function of age. Table 28 gives the multiple point biserial correlation between the dichotomized ethnic classifications and the best weighted linear composite of PPVT, Raven, and age. This multiple R is corrected for shrinkage (i.e., capitalizing on sampling error). Also shown are the partial correlations for each of the variables (PPVT, Raven, and age) with the effects of the other two partialed out. Note that the white vs. Negro partial rs are more or less equally divided between PPVT and Raven. This means that whatever is unique to each test (e.g., culture loading vs. complexity) contributes about equally to the white-Negro mean difference. The situation is quite different in the white vs. Negro comparisons. Here the major burden of the difference is attributable to the factors unique to the PPVT. The Raven factor contributes very little to the white-Mexican difference. The Negro vs. Mexican partial rs favor the Negroes on the PPVT and favor the Mexicans on the Raven.

The regression lines of PPVT on Raven and of Raven on PPVT are equally instructive. These are shown in Figure 19. The regressed

- - - - - - - - - - - - - - -

Insert Figure 19 about here

- - - - - - - - - - - - - - -

score is always shown on the Y axis. A-1 scores have been converted to standard scores (z scores). The straight arrows indicate each group's bivariate mean.

The lower half of Figure 19 shows the regression of PPVT on Raven. We see that for any given score on the Raven (the less culture loaded test), the groups' rank order from highest to lowest on the PPVT (the more culture loaded test) is white, Negro, Mexican. This is just what one might expect in predicting from a less culturally loaded test to a more culturally loaded test, especially an English vocabulary test. The upper half of Figure 19 shows the regression of Raven on PPVT. For any given score on PPVT the rank order of the three groups on the Raven, from highest to lowest, is Mexican, white, Negro. A statistical test of parallism shows that the three regression lines do not differ significantly from parallel ($F = 1.24$, $df = 4, 1654$). The intercepts of the regression lines differ significantly ($F = 52.38$, $df = 2, 1658$). And an overall test of coincidence of the regression lines shows that they differ significantly ($F = 18.30$, $df = 6, 1654$). (These statistical tests were performed on the regression lines with the effects of age partialed out.) In predicting from a more culture loaded test to a less culture loaded test, the Mexican group comes out higher than the white group, as shown in Figure 19, and this is consistent with the culture bias hypothesis

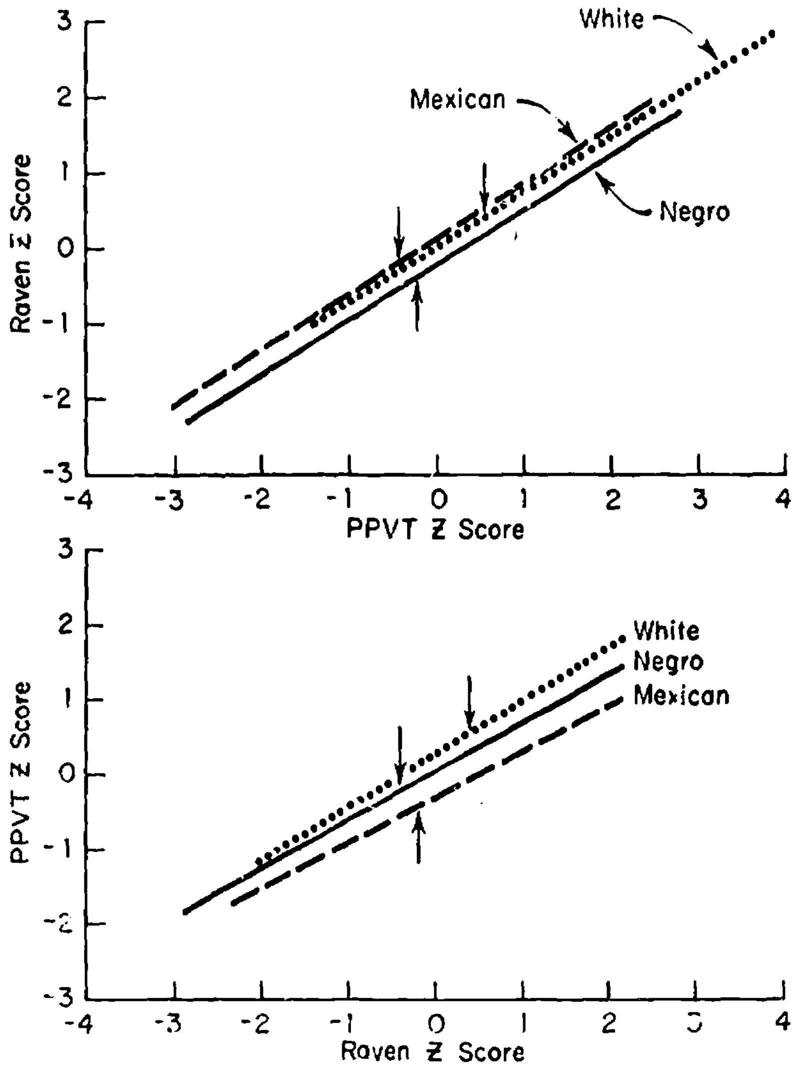



Figure 19. Regression of Raven standardized scores (z) on Peabody Picture Vocabulary Test z scores (above), and regression of PPVT scores on Raven scores (below).

of the group difference. This shows that these tests and method of analysis are capable of confirming the culture bias hypothesis of mean test score differences between groups. But the Negro group goes in just the opposite direction from the Mexican group. For any given score on the PPVT, Negroes obtain a lower score than whites and Mexicans on the Raven. The hypothesis that the white-Negro difference is mainly attributable to culture bias, in the sense in which it is defined here and is manifested by the PPVT, is therefore not supported by these data. Negro pupils do better, relative to whites and Mexicans, on the more culture loaded (PPVT) than on the less culture loaded test (Raven's Progressive Matrices). The Matrices, however, involve much more of the complexity or Level II factor than does the PPVT.

# Social Class Differences in Free Recall of Categorized and Uncategorized Lists

Arthur R. Jensen and Janet Frederiksen

In the theoretical introduction of this report it was hypothesized that Level I and Level II abilities have different age growth curves, and that the growth curves of Level I for low and middle SES are about the same, while the growth curves of Level II show an increasing divergence between the low and middle SES groups. These hypotheses are illustrated graphically in Figure 4 in the theoretical introduction.

An experimental technique that lends itself to testing this hypothesis is the free recall of categorized and uncategorized lists of familiar nouns. (These procedures are henceforth abbreviated FRC and FRU for recall of categorized and uncategorized lists.) The FRU procedure consists of showing the S a number of familiar and unrelated objects or pictures, one at a time, and after the whole list has been thus exposed, asking the S to recall as many of the items as he can remember. The same procedure is repeated for a number of trials, each time presenting the items in a different random order. The FRC procedure is the same except that the lists are composed of items which can be grouped into several conceptual categories, such as furniture, vehicles, musical instruments, etc. The single items, however, are presented in a random order on each trial without reference to their conceptual categories.

The free recall technique has two major advantages for our purposes. The first is that FRU calls primarily for Level I ability and relatively little for Level II ability, while FRC can be Level I or Level II, depending on the approach to the task that the S spontaneously chooses. FRU could conceivably engage Level II processes to a high degree, but it is much less probable that school age Ss spontaneously will bring Level II processes to bear on FRU as much as on FRC. So we can conceive of FRU as essentially a measure of Level I ability and FRC as a measure of Level II ability. The second main advantage of the free recall method, assuming that FRU and FRC do in fact measure predominantly Level I and Level II, respectively, is that there is no reason to believe that the two kinds of tests would differentially affect the Ss' motivation during the testing situation. It has been argued, for example, that intelligence tests arouse anxiety in some children, causing them to perform poorly, or that some children simply "turn off" on some tests which look too difficult or forbidding to them. A memory span test and Raven's Matrices look very different to Ss, and this difference could interact with SES, producing different favorable or unfavorable attitudinal and motivational reactions. The free recall tests, FRU and FRC, on the other hand, look alike to Ss. Everything is the same except for the fact that one list permits the items to be easily categorized. There is no reason to believe that FRU and FRC should elicit different test taking attitudes or motivational states.

Hypotheses. Two predictions can be made fr... the theory:

(1) Low and high SES groups will show a greater difference on the free recall of categorized lists (FRC) than on uncategorized lists (FRU). The basis of this prediction is that FRC involves more Level II ability than FRU, because in FRC the subject in recalling the list can transform the order of input of items to accord with the conceptual categories into which the items can be classified. The classification is a hierarchical mental process; the S notes common conceptual properties among various items, which permits classification into superordinate categories. The associations among items through their hierarchical relationship to category labels facilitates their free recall. When one item of a category is recalled, it facilitates recall of other items in the same category through their association with the common superordinate category label. Since middle SES subjects are higher in Level II ability than low SES subjects, the middle SES subjects should perform relatively better on the FRC, which can involve Level II, than on the FRU, which involves Level I. In short, the SES groups differ less on Level I than on Level II, and FRU and FRC may be regarded as tasks that typically elicit different amounts of Level I and Level II ability.

(2) The difference between lower and higher SES groups on FRC will increase with the age of the subjects. The basis for this prediction is that the hypothesized growth curves of Level II for low and middle SES groups increasingly diverge toward their different adult asymptotes as a function of age. Level II becomes an increasingly important source of individual differences and group differences variance with increasing age, going from the preschool years to adulthood.

An earlier study by Glasman (1968) tested these predictions with respect only to FRC. She used several 20-item lists of four categories each, with five items per category. The categories were: animals, foods, furniture, musical instruments, jobs, eating utensils, clothing, and vehicles. The items consisted of concrete objects -- models, toys, or other forms of real objects. The 20 items were presented singly for 3 seconds each, in a random order, for five trials. After every trial Ss were allowed 2 minutes to recall verbally the items in any order that they came to mind. The S's output was tape recorded. There were 32 Ss in each of the four groups formed by the 2 x 2 design: kindergarten vs. 5th grade and low SES vs. high SES. The low SES group was composed of Negro children from a school in a relatively poor neighborhood; the high SES group was drawn from an all white school in a middle and upper-middle class neighborhood. Thus race and SES were confounded in this study, as in the others. The mean IQs (PPVT) of the groups were 90 for low SES and 120 for high SES. The two grade levesl (grades K and 5) were matched on IQ. The main results of the study are shown in Figures 20 and 21.

- - - - - - - - - - - - - - - - - - -

Insert Figures 20 and 21 about here

- - - - - - - - - - - - - - - - - - -

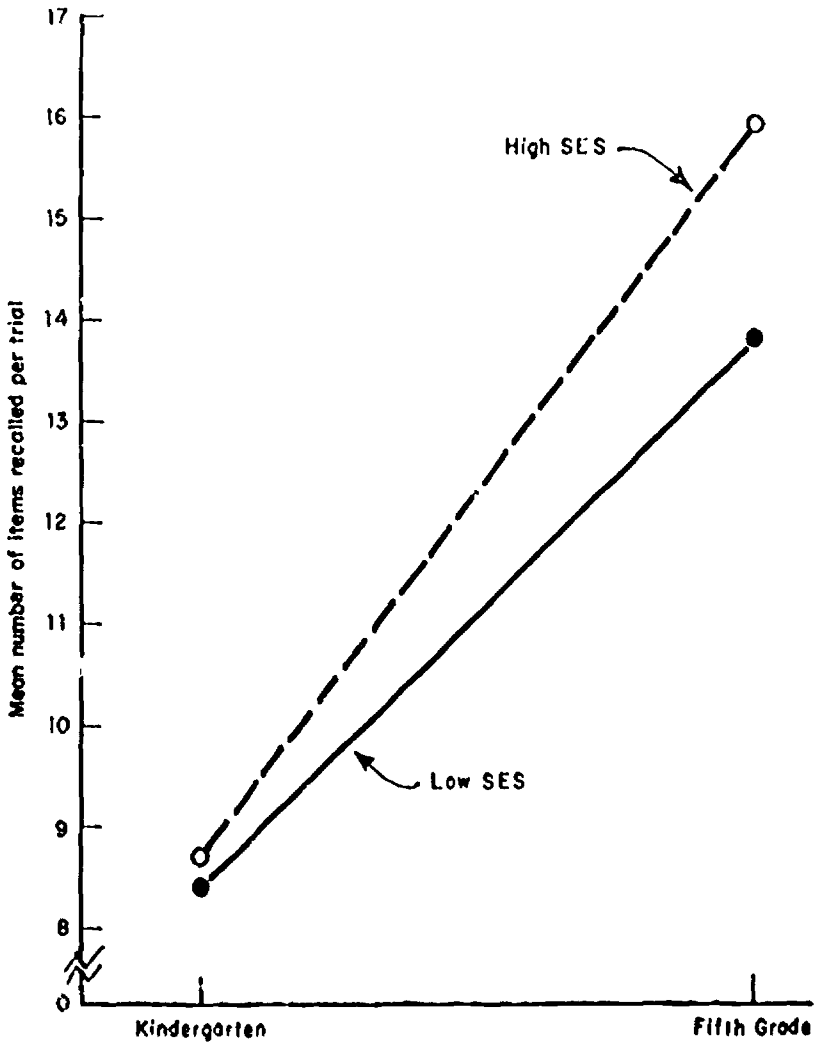



Figure 20. Mean number of items recalled per trial in free recall of categorized lists in low and high SES groups in kindergarten and fifth grade. (From Glasman, 1968).

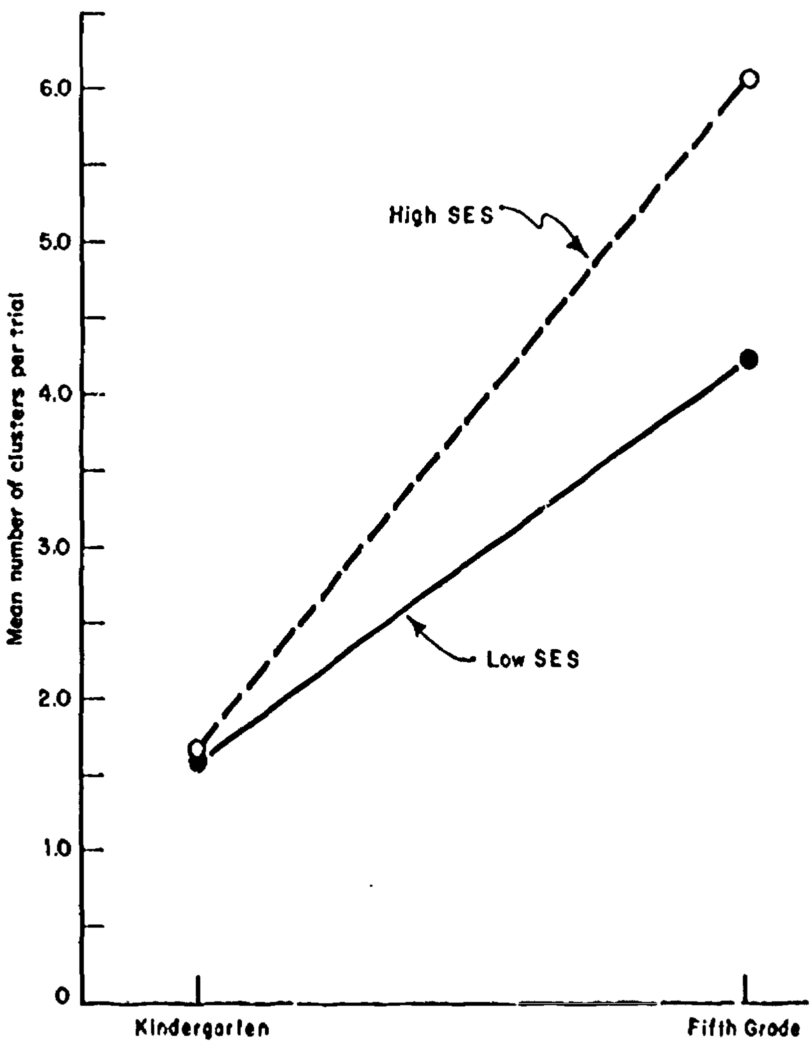



Figure 21. Clustering in free recall (Bousfield index) in low and high SES groups in kindergarten and fifth grade. Higher scores indicate greater clustering tendency. (From Glasman, 1968).

The measure of clustering used in Figure 21 is the most commonly used measure and is described by Bousfield and Bousfield (1966). A cluster is defined as a sequence of two responses from the same category which are immediately adjacent. The Bousfield formula corrects this value by subtracting the expected value for a random sequence of items recalled. The results shown in Tables 20 and 21 clearly bear out the theoretical prediction. At grade 5 the low and high SES groups differ by approximately one standard deviation, both in total amount of recall and in degree of clustering of the recall output. The grades X SES interaction is statistically significant beyond the .05 level for recall and beyond the .001 level for clustering.

Since FRC elicits Level II processes, it should be correlated with mental age in both low and high SES groups. This is what Glasman found. The correlation between MA and amount of recall was .62 for the low SES and .72 for the high SES group. The correlation between MA and the amount of clustering was .76 for low SES and .77 for high SES. The correlations are much higher for 5th graders than for kindergartners, who show very little clustering and are presumably operating in this task by a Level I process. The correlation of MA and recall is .06 at kindergarten and .59 at grade 5. The correlation between MA and clustering is .02 at kindergarten and .68 at grade 5. FRC performance is so strongly related to MA that when the data of Figure 20 and 21 were subjected to an analysis of covariance with MA as the control variable, all the main effects and the interactions were completely wiped out. It thus appears that the FRC task is a kind of IQ test and probably correlates as highly with standard IQ tests as the reliability of the FRC scores (recall and/or clustering) will permit, at least for children in the 5th grade. This fact gives an interesting insight into the nature of Level II ability.

Although Glasman's study demonstrated age and SES differences in the free recall of categorized lists, it was not designed to study age and SES differences in categorized versus uncategorized lists. An uncategorized list is made up of unrelated or very remotely associated items which cannot be readily grouped according to superordinate categories. Subjective organization of the items in the list is most likely to consist of pairs of items related on the basis of primary stimulus generalization, clang association, or functional relationship. An uncategorized list therefore lends itself less than a categorized list to evoking Level II processes. Consequently, subjects differing in Level II ability (but not in Level I) should show less difference in FRU than in FRC. The present experiment was intended to test this prediction.

## Method

### Subjects

Negro and white 2nd and 4th grade children, 120 in all, were selected from two schools, one in a low SES neighborhood and one in a middle to upper-middle class neighborhood. The groups were very similar in composition to those used in Glasman's study. Ten children in each grade within each school were randomly assigned to one of three experimental conditions: an uncategorized list (U), a random categorized list (RC), and a blocked categorized (BC) list. It is thus a 2 x 2 x 3 factor design, the factors being SES (Negro-low vs. white-high), Grades (2nd vs. 4th) and Lists (uncategorized vs. categorized-random vs. categorized-blocked).

## Procedure

Ss were tested individually. Each was presented with a set of 20 familiar objects and was told he would have to remember and recall the names of all the objects he was shown. The objects were presented serially. The uncategorized list consisted of the following toy objects: ball, bell, book, box, brush, car, chair, clock, coat, cup, egg, flag, frog, gun, horse, key, pen, thread, train, wheel. The items were presented in a different random order on each trial. The categorized list consisted of items representing four categories: clothing, tableware, furniture, and animals. The items were: coat, dress, hat, shoe, skirt, cup, glass, plate, spoon, knife, mouse, chicken, dog, horse, cow, bed, chair, dresser, lamp, table. The items were presented in a different random order on each trial. The categorized-blocked list consisted of the same items but all the items of one category were always presented in sequence. The items were presented in a different random order within category blocks on each trial, and the order of the category blocks was varied randomly on every trial.

Each S was given five learning-recall trials on one of the three sets of objects. As each object was presented, the S was asked to name it. E accepted the S's name for the object or provided the name if S did not respond. Virtually all Ss could name all the objects without hesitation. Each object was removed from view before the next was presented. The rate of presentation was approximately 2 seconds per item. When all 20 items had been seen and named by the subject, he was given 90 seconds to recite the names of all the objects that he could recall. This procedure was repeated for five trials. Instructions to the subjects and all other features of the testing procedure were exactly the same for the three lists -- U, RC, and BC. It should be clearly understood that no S was tested in more than one of the experimental conditions.

E recorded S's responses and their order of emission on a specially prepared form. All Ss were tested by Mrs. Frederiksen.

# Results

## Amount of Recall

The recall measure was number of correct responses over five trials. The results for the three experimental treatments, Uncategorized (U), Categorized (C) and Blocked (B), are shown in Figures 22, 23, and 24. These figures are interpretable in connection with

- - - - - - - - - - - - - - - - - - - - - - - -

Insert Figures 22, 23, and 24 about here

- - - - - - - - - - - - - - - - - - - - - - - -

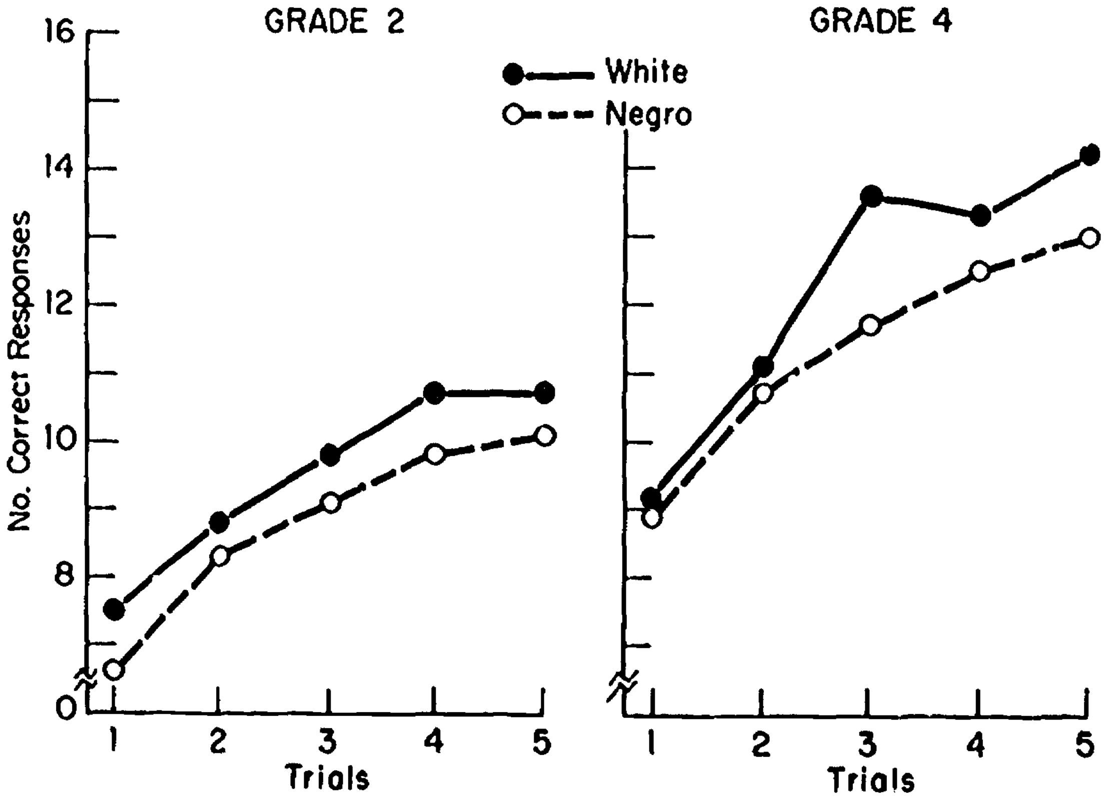



Figure 22. Amount of free recall of random uncategorized list.

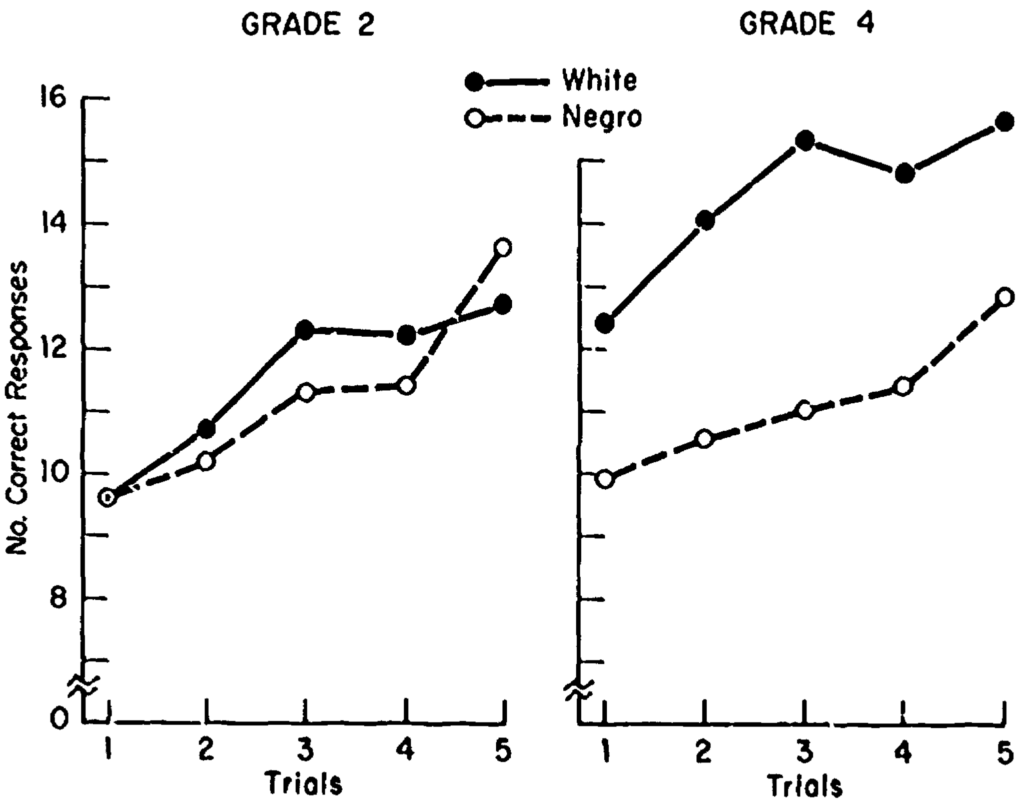



Figure 23. Amount of free recall of random categorized list.

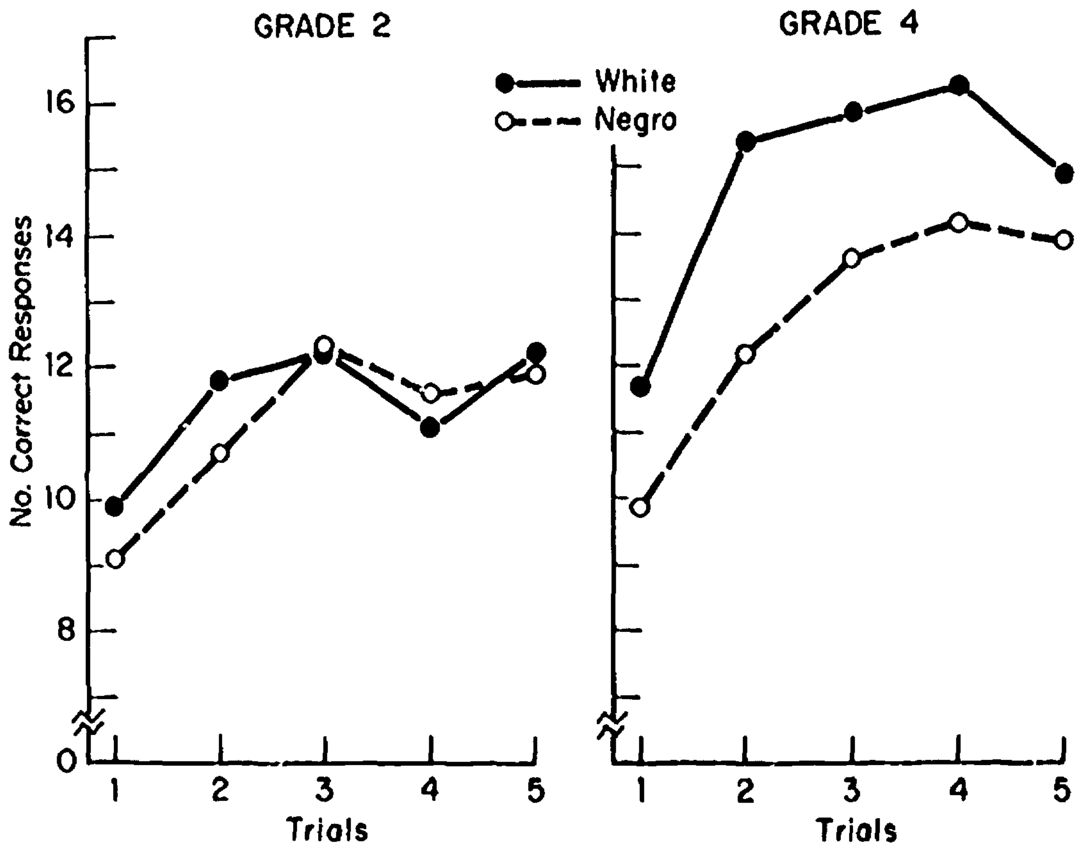



Figure 24. Amount of free recall of blocked categorized list.

the analysis of variance. A multivariate analysis of variance was used in which the five trials are treated as a mean vector. The mean vectors are tested for the statistical significance of differences between groups in a nested design. Race (R) is nested in Grades (G) and Treatments (T), and Grades are nested in Treatments. The rationale and methodology of this design, which is most appropriate for the analysis of the present experiment, has been fully explicated by Marascuilo and Levin (1970). The analysis is summarized in Table 29.

- - - - - - - - - - - - - - - -

Insert Table 29 about here

- - - - - - - - - - - - - - - -

The analysis shows that overall the Treatments (Uncategorized, Categorized, and Blocked lists) differ significantly.

With reference to Figure 22, the analysis shows that the overall difference between grades 2 and 4 is significant ($p < .001$). The Negro and white groups, however, do not differ significantly in either grade level. This is in accord with the hypothesis that the Uncategorized list is essentially a Level I learning task which should show little difference between lower and upper SES groups. The significant grade difference reflects the growth of Level I ability during this age period.

In Figure 23, the grades do not differ significantly. The white vs. Negro difference is not significant in grade 2 but is significant ($p < .014$) in grade 4. This accords with the hypothesis that the Level II ability (evoked by the Categorized list) has a steeper growth curve in upper than in lower SES Ss, as represented here by white and Negro groups, respectively. At Grade 2 (approximately age 7) the groups are not very differentiated in Level II ability, at least as it is evoked by this task.

In Figure 24, the grades differ significantly ($p < .003$). The racial (SES) groups, however, do not differ significantly. There was no prior hypothesis about this condition. It was included to find out if making the categories more obvious by blocking would facilitate clustering and recall in Ss for whom a random categorized list does not evoke Level II processes. It appears that both racial (SES) groups are facilitated by blocking, the Negro more so than the white, so that the groups do not differ significantly under this condition.

Category Clustering

Ss' clustering of their free recall in the Categorized and Blocked lists was measured by means of a clustering index, Z, which was devised as an improvement over other measures of clustering, all of which present certain problems that are overcome by the Z index (Frankel & Cole, in press). The Z index is based on the statistical properties of runs. A run is defined as a number of items from the same category that are recalled successively. The length of each run is the number of successive items from the same category. Single items are regarded as runs of one. The expected mean ($EM_r$) and variance ($EV_r$) for the

Table 29

Multivariate Analysis of Variance (Nested Design) for Free Recall:

Number of Correct Responses

(As Represented by a Mean Vector for Five Trials)

| Source of Variance | df | F | p |
|---|---|---|---|
| Treatments (T) | 2 | 7.36 | <.001* |
| Grades (G) in Treatments | (3) | | |
| G in $T_{Uncategorized}$ | 1 | 4.57 | <.001* |
| G in $T_{Categorized}$ | 1 | 1.54 | <.185 |
| G in $T_{Blocked}$ | 1 | 3.91 | <.003* |
| Race (R) in Grades (G) and Treatment (T) | (6) | | |
| R in $G_2$ $T_U$ | 1 | <1 | <.962 |
| R in $G_4$ $T_U$ | 1 | 1.62 | <.162 |
| R in $G_2$ $T_C$ | 1 | 1.32 | <.261 |
| R in $G_4$ $T_C$ | 1 | 3.03 | <.014* |
| R in $G_2$ $T_B$ | 1 | <1 | <.616 |
| R in $G_4$ $T_B$ | 1 | 1.42 | <.223 |
| Error | 108 | | |

*Significant effects, p <.02

number of runs in a randomly selected list of arbitrary length N and number of categories C can be statistically computed (Wallis & Roberts, 1956, p. 571). The Z index of clustering is

$$Z = \frac{EM_r - O_r}{\sqrt{EV_r}}$$

where $O_r$ is the observed runs

$EM_r$ is the expected mean runs in a random series of the same length (N) and number of categories (C) as the observed recall series

$\sqrt{EV_r}$ is the expected standard deviation of runs in a random series with N and C the same as in the observed series.

The Z is thus a standard score referable to the table of the normal distribution for its probability of occurrence. Clustering is defined as the presence of significantly "too few" runs, i.e., fewer than would occur in a random output of the same items. As can be seen from the above formula, larger Z scores indicate a greater degree of clustering. It is a pure measure of clustering, independent of amount recalled.

Figures 25 and 26 show the group results for the clustering Z scores.[1] The method of statistical analysis is the same as that used

- - - - - - - - - - - - - - - - - - - -
Insert Figures 25 and 26 about here
- - - - - - - - - - - - - - - - - - - -

for the recall data; it is summarized in Table 30. With regard to

- - - - - - - - - - - - - - -
Insert Table 30 about here
- - - - - - - - - - - - - - -

Figure 25, the analysis of variance shows no significant overall grade difference in clustering of the Categorized list. The white vs. Negro difference is not significant at grade 2 but is significant ($p < .005$) at grade 4. This is in accord with our hypothesis that Level II is reflected in clustering (i.e., conceptual transformation of input prior to output) and that it has a steeper growth function in high (white) than in low (Negro) SES groups.

As to Figure 26, the Blocked condition, the analysis indicates a significant grade difference. Clustering tendency is evoked by blocking in more 4th than 2nd graders. The racial difference in clustering is not significant.

---

[1] We are indebted to Dr. Michael Cole, Rockefeller University, for obtaining all the Z scores from our data by means of a computer program he has devised for this purpose.

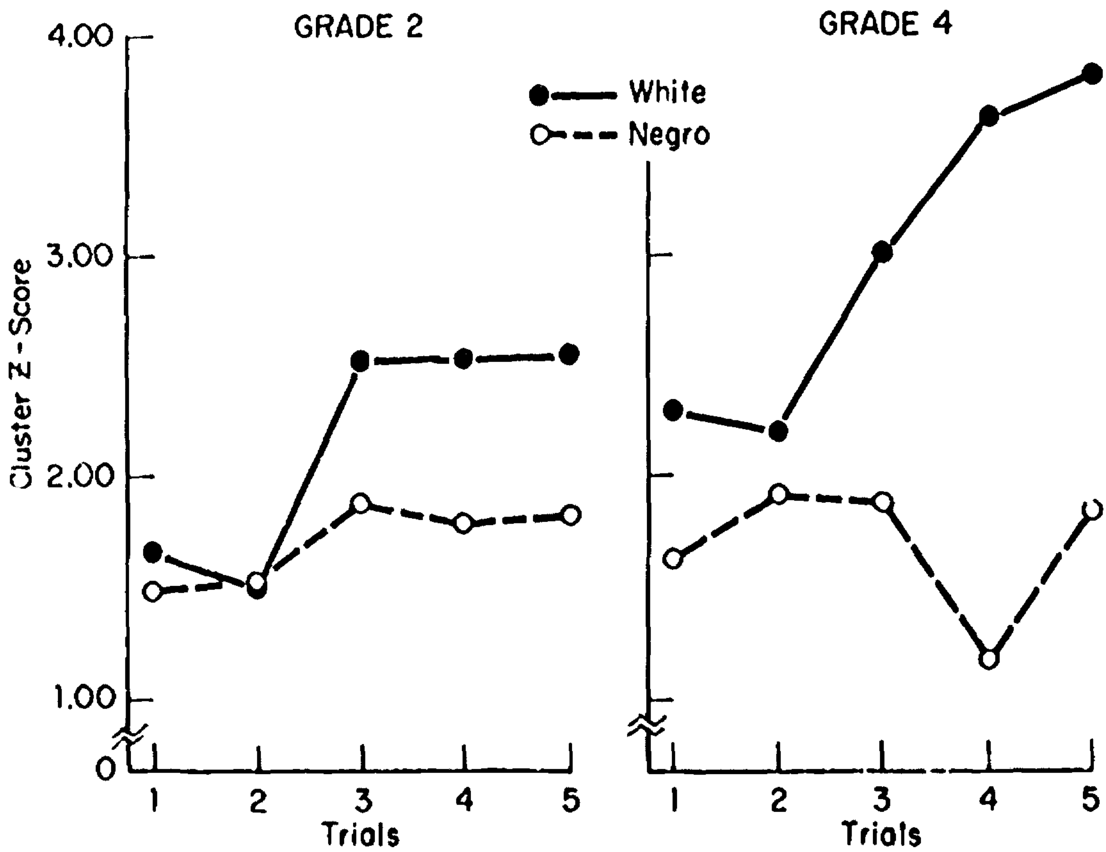



Figure 25. Amount of clustering in free recall of random categorized list.

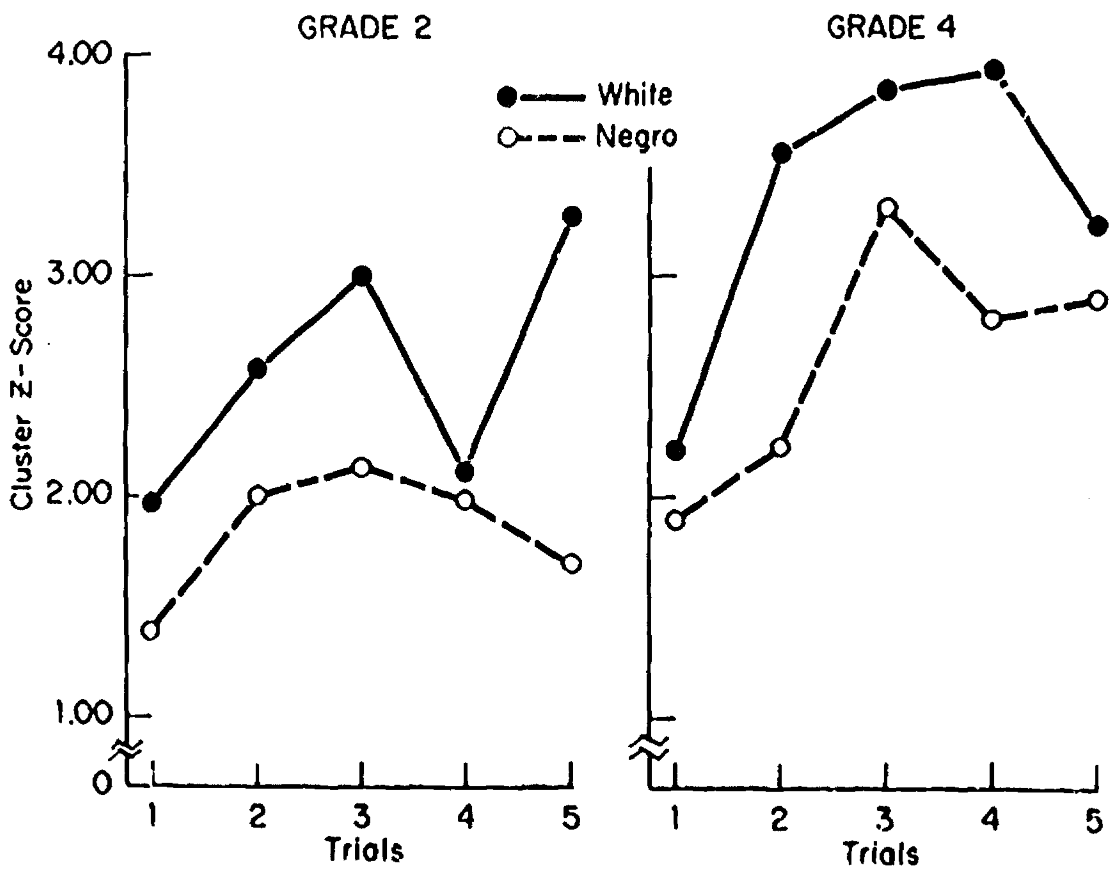



Figure 26. Amount of clustering in free recall of blocked categorized list.

Table 30

Multivariate Analysis of Variance (Nested Design) for Free Recall: Clustering _Z_ Score (As Represented by a Mean Vector of Five Trials)

| Source of Variance | df | F | p |
|---|---|---|---|
| Treatments (T) | 1 | 2.75 | <.026* |
| Grades (G) in Treatments | (2) | | |
| G in $T_{Categorized}$ | 1 | <1 | <.686 |
| G in $T_{Blocked}$ | 1 | 2.91 | <.019* |
| Race (R) in Grades (G) and Treatments (T) | (4) | | |
| R in $G_2$ $T_C$ | 1 | <1 | <.680 |
| R in $G_4$ $T_C$ | 1 | 3.75 | <.005* |
| R in $G_2$ $T_B$ | 1 | 1.40 | <.235 |
| R in $G_4$ $T_B$ | 1 | 1.52 | <.195 |
| Error | 72 | | |

*Significant effects, p <.03

Summary

It was hypothesized that low SES (Negro) and middle SES (white) groups would differ less in the free recall of random than of categorized lists and that the difference would be greater in older than in younger children. Both hypotheses were borne out by the data. It was also hypothesized that clustering tendency in the recall of categorized lists would be greater in the high than in the low SES group and that the difference would be greater for older children. These hypotheses were also substantiated.

# Mental Elaboration and Learning Proficiency[1]

William D. Rohwer, Jr.

Individual differences in learning proficiency are exemplified in a wide variety of phenomena. Interest in two such phenomena has spawned the work to be reported in the following two papers. One of these phenomena may be described as individual differences in performances that convey information; the other can be characterized as differences across persons in their capacity to juxtapose effectively disparate kinds of information. This phenomenon is more difficult than the first to describe in rigorous or even quasi-rigorous terms but the effort is clearly called for since the present description is murky indeed.

One way to think about the subject is in terms of a contrast between the manner in which information is conveyed in a poem and the manner in which it is conveyed in a tightly-reasoned logical argument. The predictability of expository presentation is relatively low and yet the internal congruence of the two kinds of presentation may be equivalent. In expository writing, the substantive content is often so well-organized that it can almost be described by a formal set of rules, whereas in poetry, the content or substance of the message does not yield to logical enumeration; comprehension requires imaginative rather than formal conceptual activity.

With this contrast in mind, start with a first assumption: efficient, successful learning necessarily involves conceptual, in contrast to rote, processing of information. Such conceptual activity can vary in character along a dimension that stretches from the pole of formal processing on one end to the pole of imaginative processing on the other. The presumption is that the acquisition of information, like the presentation of information, is either formal-dominant or imaginative-dominant; if you will, it is either logical or poetic. Let there be no mistake: in both cases, formal activity or imaginative activity, information is organized -- it is the manner of its organization that differs.

To come down to the earth of experimental psychology, the contrast can be illustrated in connection with a well-known methodological variation in research on free recall. If one is interested in organizational activities engaged in by subjects performing on free-recall tasks, he can proceed in one of two ways. He can use a list of stimulus items selected from well-defined classes and observe the imprint of these classes on the order or the amount of items recalled. For example, such a list might be comprised of names of four seasons, names of four directions, names of four animals and names of four vehicles. In this method, the method of categorized lists, the focus is on the utility of a formal system in fostering the acquisition of a set of items. The alternative is the method of uncategorized lists wherein stimulus items are selected such that no two of them are drawn from the same class. In this case, the investigator inspects the subject's response output for evidence of self-generated organization which is usually not characterized by the use of a formal system but by idio-

---

[1]This paper is an abridged version of a chapter bearing the same title that appears in Hill, J.P. (Ed.), Minnesota Symposia on Child Psychology, Volume IV. Minneapolis: University of Minnesota Press, 1970.

syncratic or subjective or imaginative ways of grouping items. The one method tends to elicit formal conceptual activity during acquisition while the other tends to elicit imaginative conceptual activity.

Of these two sorts of conceptual activity our concern has been with the imaginative side of the dimension, that is, with the effects of imaginative mental activity on learning efficiency. Elsewhere I have denoted the topic of this concern with the terms mental mnemonics (Rohwer, 1968) mnemonic elaboration and mental elaboration (Rohwer, 1967). The last of these is preferable to the first two since it does not prejudge the issue whether the conceptual activity referred to does, in fact, facilitate learning. The meaning of the word mental is already as clear, no doubt, as it can be made (which is not to say that it is very clear) but elaboration deserves additional explication.

The need for some such word can be appreciated readily by reflecting on the fact that imaginative conceptual activity during learning can proceed in the direction of selecting for attention only parts of the materials presented, that is, by reducing the amount to be acquired (e.g., stimulus selection) or it can proceed in the direction of augmenting the materials presented, that is by elaborating on the elements to be acquired. An example of elaborative activity is provided by a manipulation that can be performed in an experiment using the method of paired-associates (PA) learning. Suppose the task is to learn a list of noun pairs in such a way that when one member of each pair is presented, the other member can be recalled. Before the pairs are initially presented, subjects can be given one of two kinds of instructions: to read aloud the nouns as they appear, or to construct and utter a sentence containing the two nouns as they appear. In complying with the sentence instructions, a subject is engaged in mental elaboration, that is, he is elaborating the noun pairs into sentences.

The only benefit to be derived from the discussion to this point is that it permits a restatement of the major phenomena that have initialed and guided our program of research: (1) the effects of mental elaboration on learning efficiency; and (2) individual differences in learning proficiency, especially as they arise from individual differences in mental elaboration.

A research effort directed toward this goal of increasing our understanding of these phenomena commits itself to work on two major tasks: that of identifying and subjecting to experimental analysis those forms of mental elaboration that are successful in increasing learning efficiency; and, that of determining whether or not a specification of these forms of mental elaboration provides any assistance in understanding the differences between more and less proficient learners. Interest cannot remain confined solely to these tasks, however, since in the course of work on them, a number of other substantive issues are raised which also command attention. Among these other issues are: the role of imagery in learning; the role of language in learning; the notion of mediation; the developmental primacy of imaginal and verbal processes in learning; stimulus conditions and learning efficiency; the developmental theory of mental retardation; ethnic and socioeconomic differences in learning and intelligence;

predicting school success; and, diagnosing strengths and weaknesses in elaborative skills.

## Mental Elaboration

A variety of methods for conducting elaboration research are readily available in the rapidly increasing literature on the topic. At least three major ones may be distinguished: the method of post-learning interviews; the method of instructions; and the method of manipulating stimulus conditions. Each of these can be applied with equivalent facility to the PA paradigm, the serial paradigm and the free-recall paradigm, and each has advantages and disadvantages for this purpose.

The method of post-learning interviews has a special appeal since it seems better suited than either of the other methods to the purpose of revealing directly the character of the subjects' own elaborative activities rather than those of the experimenter. In brief, the method consists of presenting a list of PAs for learning and at some point, either during or after acquisition is complete, asking the subjects to describe for each pair the technique they used to remember it. Such interviews do elicit reports of conceptual activity (Bugelski, 1962; Runquist & Farley, 1964) and these activities can be classified with respect to their complexity (Martin, Boersma & Cox, 1965). The most complex category clearly falls within the domain of elaborative activity as its typical expression is the formation of sentences containing the two members of a pair. Interestingly enough, further studies have revealed that the more complex the mental activity, the better the learning so that the most efficient acquisition is associated with elaboration (Martin, Cox, & Boersma, 1965; Montague & Wearing, 1967). Results of this sort, obtained by the post-interview method, lend credence to the notion that learning is accompanied by conceptual activity and that efficient learning is associated with elaborative activity.

Despite its directness, however, the post-interview method leaves several questions entirely unanswered. Foremost among these is whether elaborative activity is responsible for efficient learning or only an epiphenomenal accompaniment of it. For example, is a PA elaborated and therefore learned or is the PA learned and elaborated afterwards? Other questions concern the accuracy with which subjects can characterize the conceptual activities in which they engage during learning and the problem of isolating those aspects of reported elaboration that are responsible for increases in learning efficiency as against other aspects that are extraneous to such increases.

With respect to these issues, the second method, that of instructional manipulation, has distinct advantages over the method of post-learning interviews. In its simplest form, this method generates a two-group experiment: one group is instructed to elaborate each noun pair in the PA list as it is presented whereas the other group is simply instructed to attend to and remember the noun pairs. Two forms of elaboration instructions have been used -- sentence instructions and imagery instructions. Both forms produce remarkable amounts of facilitation in PA learning among college students although there is some ambiguity about whether imagery instructions produce more facilitation than sentence instructions (Bower, 1967, 1969) or equivalent amounts (Paivio & Yuille, 1967). The efficacy of sentence instructions has also been established for samples of children and of retardates (Jensen & Rohwer, 1963, 1965: Milgram, 1967a).

The results of experiments that involve the manipulation of elaboration instructions permit a stronger inference than do the post-learning interview experiments: that elaboration does increase learning efficiency. That is to say, if pairs are elaborated upon during initial presentation, they are learned more rapidly than if they are not elaborated. In addition, however, the results present two problems. First, they suggest that subjects do not habitually and systematically engage in the forms of elaboration prompted by such instructions in typical PA learning experiments; otherwise facilitation relative to an ordinary control condition would not have been observed. Still, this is not to say that subjects never engage in such activities spontaneously since instructions that interfere with their opportunity to do so (e.g., instructions to rehearse each noun pair) have the effect of depressing performance (Bower, 1969). Moreover, post-interview studies have revealed that the amount of elaborative activity engaged in by a single subject varies considerably across a list of PAs. The second problem presented by these results is that of specifying the properties of the elaborative activities elicited by instructions that are necessary and sufficient for the facilitation of learning.

This second problem highlights the strength of the third method that has been used for investigating the effects of elaboration on learning, the method of manipulating stimulus and response conditions. This strength is that the properties of both verbal and visual forms of elaboration are under the control of the experimenter and, therefore, can be varied systematically to assess their effects upon learning efficiency. The method has a glaring weakness as well -- the degree to which externally presented elaboration corresponds to internal elaborative activity remains entirely unknown as does the character of the conceptual processes prompted by experimenter-controlled elaboration. Doubts about these issues are partially allayed by the results of the other two methods of conducting elaboration research since they converge on the conclusion that subjects do indeed use many of the forms of activity that have been manipulated externally. Accordingly, the potential of the method outweighs its disadvantages sufficiently to warrant using it and the succeeding discussion describes some of its yield.

Verbal Elaboration

Except for a few experiments reported by Epstein, Rock and Zuckerman (1960), the experimental analysis of verbal elaboration began with a study of noun-pair learning in sixth-grade children (Rohwer, 1966). Starting with the fact that the presentation of noun pairs in sentence contexts facilitates acquisition (Jensen & Rohwer, 1963) the experiment was designed to determine whether or not the sentence unit was a necessary condition for such facilitation. Accordingly, the noun pairs were presented in three different contexts distinguished by the form class of the word that linked the two members of each noun pair. With the exception of this linking word or connective, the number and identity of all the words in the three kinds of context were the same. The three connectives were conjunctions, prepositions and verbs. By way of illustration, here are the three contexts for the pair, COW-BALL:

Conjunction: The running COW <u>and</u> the bouncing BALL.
Preposition: The running COW <u>behind</u> the bouncing BALL.
Verb: The running COW <u>chases</u> the bouncing BALL.

In Figure 27, the percentages of correct responses per trial are plotted as a function of connective form class. Two control

---

Insert Figure 27 about here

---

conditions were used: an ordinary PA condition in which the noun pairs were presented without context; and a consonant control (CC) where the nouns were presented in the context of a string of consonants - b f COW x m d BALL. Relative to the PA control, both the verb and preposition connectives produced facilitation; the conjunction did not. The difference between the verb and preposition groups was not significant but performance in the PA control, being indistinguishable from the conjunction condition, was superior to the CC condition. Apparently, the consonant strings interfered with whatever autonomous learning activity the subjects engage in under ordinary conditions of PA learning.

The specific phenomenon revealed in this experiment, that is, the form-class effect, demonstrates that only particular kinds of verbal elaboration promote efficient learning; facilitation does not occur irrespective of the kind of elaboration used. Consequently, the form-class effect has prompted a number of other investigations in an effort to give a general account of the features of verbal elaborative activity that are necessary for facilitation. Most of these studies have been designed to examine the role of certain linguistic variables, both syntactic and semantic, in verbal elaboration while the remainder have concerned the impact of selected task variables on the form-class effect (Ehri & Rohwer, 1969; Jensen & Rohwer, 1963, 1965; Paivio, 1967: Paivio & Yuille, 1967; Rohwer, 1966; Rohwer & Ammon, 1968; Rohwer & Levin, 1968; Rohwer & Lynch, 1967; Rohwer, Lynch, Levin & Suzuki, 1967; Rohwer, Shuell & Levin, 1967; Suzuki & Rohwer, 1968; Suzuki & Rohwer, 1969).

<u>Visual Elaboration</u>

Recall that experiments using the method of elaboration instructions have shown that visual, or imagery, instructions facilitate learning as well as verbal, or sentence, instructions (Bower, 1967, 1969; Paivio & Yuille, 1967). Accordingly, the phenomena of visual

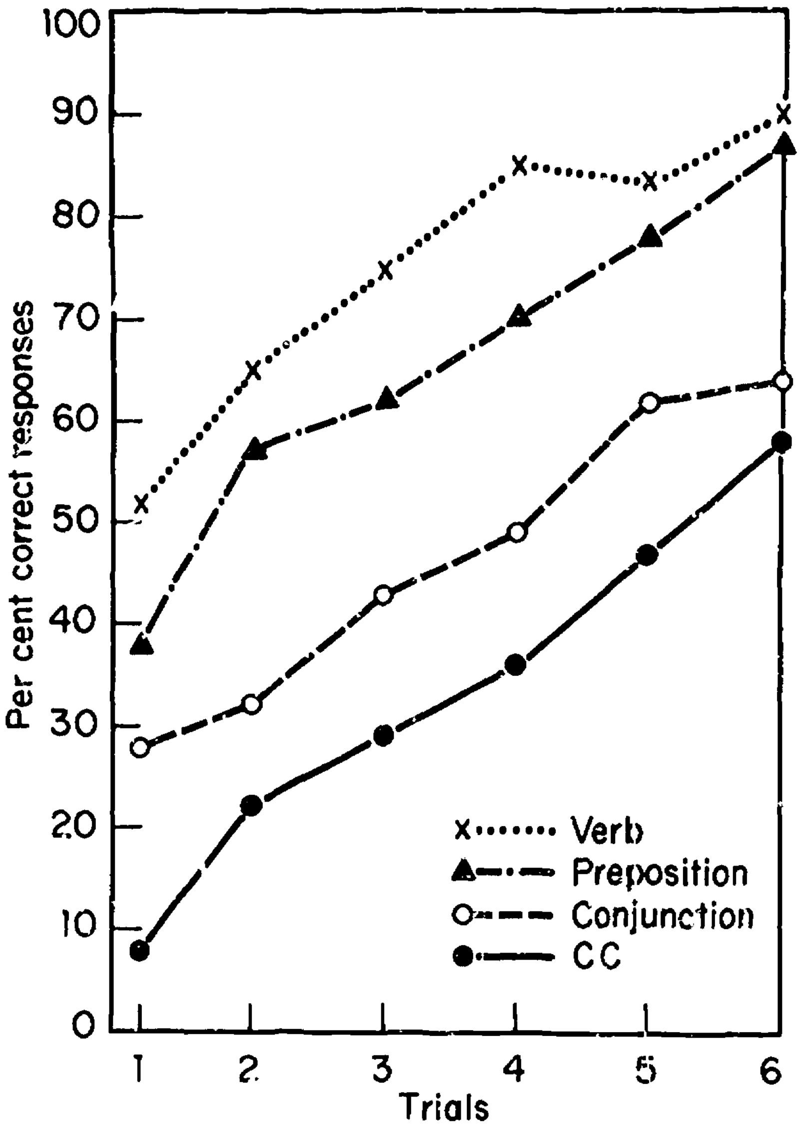



Figure 27. Mean percentages of responses correct per trial as a function of connective form class.

elaboration and efficient learning have provoked attempts at experimental analysis similar to those made for the phenomena of verbal elaboration. Indeed, in one experiment, the intent was to vary verbal and visual elaboration in direct parallel (Rohwer, Lynch, Suzuki & Levin, 1967). For each of 24 noun pairs, three kinds of contextual verbal strings were created -- conjunction strings, preposition strings and verb strings. Then, for each of the three lists of contextual strings, a corresponding list of pictorial materials (recorded on movie film) were constructed in such a way as to constitute a visual translation of the objects, situations and events described by the contextual strings. That is to say, the pairs of objects named by the nouns were photographed in three different ways: (1) Still (conjunction) -- every pair of objects was placed on a table and photographed; (2) Locational (preposition) -- the pairs of objects were placed on the table in a way depicting a particular spatial relationship between them (e.g., one object inside, above, behind, beneath the other); and, (3) Action (verb) -- the objects in every pair were photographed while in motion, depicting some kind of action episode. By way of illustration, consider the materials for the pair DOG-GATE. In the Still condition, the subject would simply see a picture of a dog and a gate. In the Locational condition, the picture would show a dog perched on top of a gate. And in the Action condition the picture would show the dog literally walking to the gate and closing it.

Each of the pictorial or depiction conditions was presented under four different conditions of verbalization: Naming, Conjunction, Preposition and Verb. All of these materials were administered to samples of first-, third- and sixth-grade children. There were no significant interactions between conditions, effects and grade level, so the results presented in Figure 28 represent performance averaged across the samples.

Insert Figure 28 about here

One of the most interesting aspects of these results is that the effect associated with the Depiction variable is quite similar to the effect associated with the verbalization variable, that is, the form-class effect. This outcome suggests the possibility that the process underlying the form-class effect might be visual in nature, that is, it might involve imagery. The matter is not at all clear, however, since the results also suggest the opposite possibility, namely that a covert verbalization process may underlie the depiction effect. None of the numerous attempts reported thus far to settle the issue empirically has allowed for a conclusive choice among these two possibilities (Reese, 1965, 1970; Milgram, 1967b; Paivio, 1970; Palermo, 1970; Rohwer, 1967, 1970).

One promising way of attacking the issue is to phrase the question developmentally. Assume that older children and adults have available at least two ways of representing information in memory, verbal and visual. Then one of the questions that may be asked is: In connection

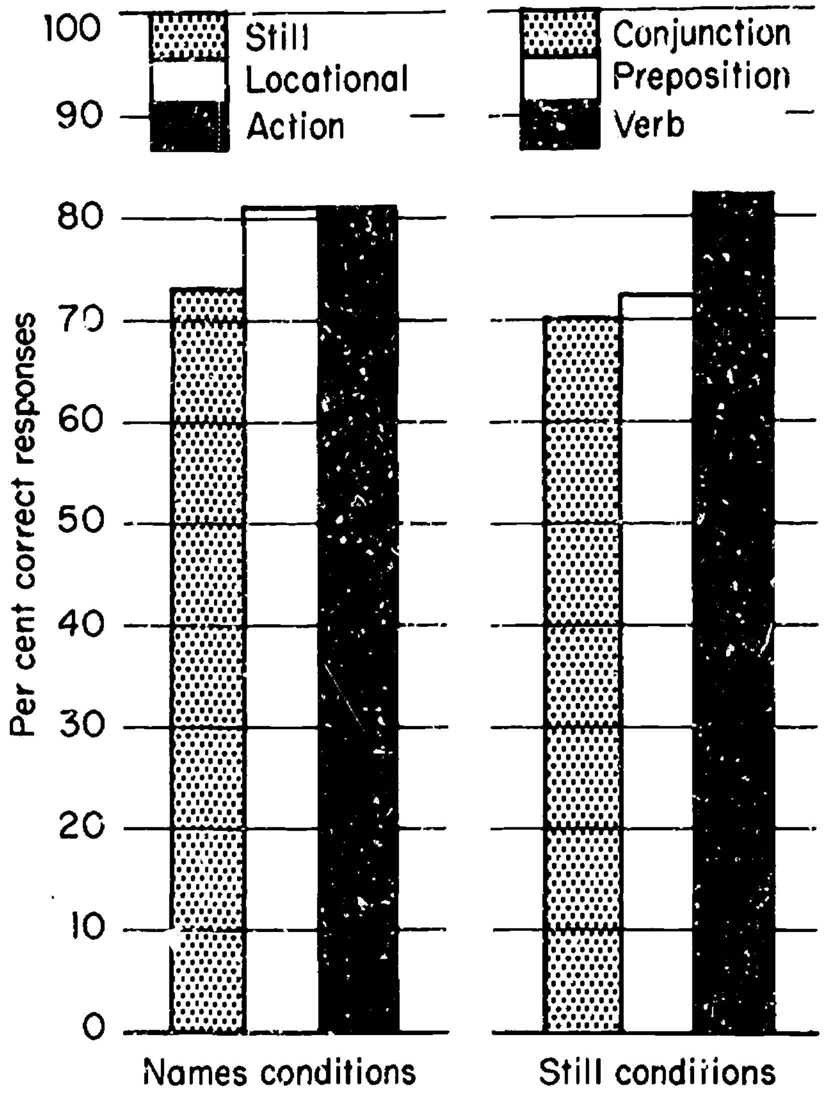



Figure 28. Mean percentages of responses correct for Names conditions as a function of depiction and for Still conditions as a function of connective form class.

with children's learning, does the visual or the verbal mode of memory storage emerge earlier developmentally? It can be argued that one method for determining whether visual or verbal modes of storage are dominant is to assume that the dominant mode will be associated with more efficient learning than the less dominant mode. If learning efficiency is accepted as an index of the dominance of one or the other of the two modes, then the developmental issue can be investigated in a similar fashion, that is, data can be examined for evidence that the effect on learning efficiency of mode differences varies as a function of age.

Elsewhere, I have advanced in detail the hypotheses that (1) the visual mode is generally dominant and (2) the degree to which the visual mode is dominant over the verbal <u>increases</u> with age (Rohwer, 1970). The first hypothesis is consistent with data such as those reported by Paivio (1967), to the effect that high-imagery words are easier to learn than low-imagery words, by Dilley and Paivio (1968), showing that pictures produce better learning than words, and by Rohwer, Lynch, Levin and Suzuki (1967), also showing that more efficient learning is associated with pictures than with words.

The second hypothesis runs counter to the widely disseminated notion that pictorial or iconic modes of representation are developmentally more primitive than verbal modes of representation (cf. Bruner, 1966). Nevertheless, there are data to support this hypothesis from experimental studies of learning in children. In one study, for example, four mixed lists of 25 noun pairs were administered to samples of kindergarten, first- and third-grade children by means of videotape played through a television monitor (Rohwer, 1969). The lists were mixed with respect to the five different ways in which the pairs were presented: Names -- nouns presented aurally without visual depiction; Still -- pictures of object pairs without aural naming; Names-Still -- a combination condition with pictures of objects and their noun names presented aurally; Sentence-Still -- pictures of object pairs with a sentence containing their noun names presented aurally; and Names-Action -- action pictures of object pairs with their noun names presented aurally. Every list consisted of five pairs of each of these five types presented in a random order.

For all three grade levels the order of the pair types with respect to the associated degree of learning efficiency (from least to most) was: Name, Still, Name-Still, Sentence-Still, Name-Action. The results for the first two of these pair types are pertinent for a test of the hypothesis that the dominance of the visual over the verbal mode increases with age. The difference between the Still items and the Name items is plotted in the upper panel of Figure 29 as a function of age. Note that the superiority of pictorial items

- - - - - - - - - - - - - - -
Insert Figure 29 about here
- - - - - - - - - - - - - - -

over verbal items increases with age, as predicted. A similar trend is apparent in related data reported by Dilley and Paivio (1968).

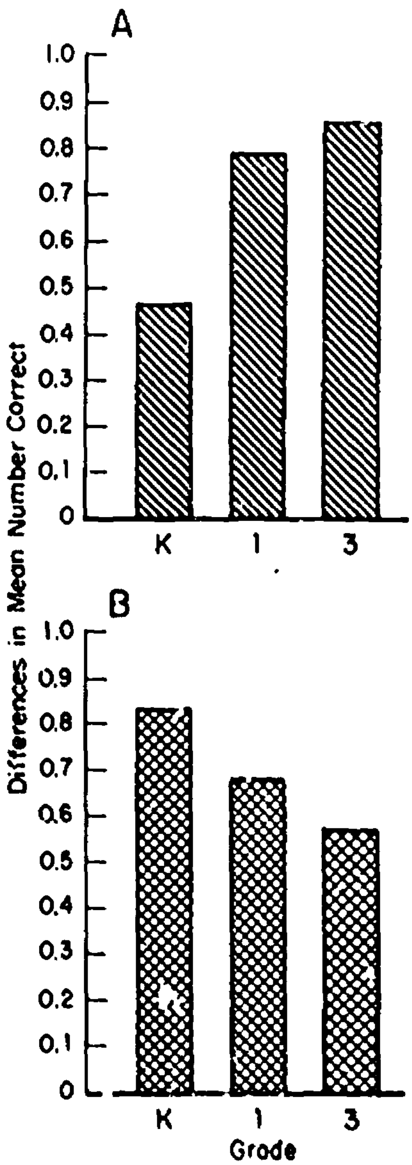



Figure 29. (A) Differences between mean performance on Still and Names items as a function of grade level. (B) Differences between mean performance on Names-Still and Still items as a function of grade level. (From W. D. Rohwer, Jr., Images and pictures in children's learning: Research results and educational implications. Psychological Bulletin, 1970, 73, 393-403.)

If the hypothesis is correct that the general superiority of the visual over the verbal mode of storage increases with age, the problem is to account for the effect. One explanation is that the efficacy of visual storage is amplified when complementary verbal representation is stored simultaneously with the visual. The corollary developmental assumptions, of course, are: (1) that the capacity of children for the simultaneous storage of information in two modes increases with age, especially across the interval stretching from three- or four- to seven- or eight-years-old; and (2) that across this same interval, the child's capacity for generating a verbal representation of an object or event also increases. This explanation is consistent with the overall superiority of the Name-Still over the Still picture conditions and the first assumption is consistent with the significant increase in performance in both conditions as a function of age. The second assumption, however, requires a different sort of confirmatory evidence; specifically, it makes the prediction that the relative superiority of the Name-Still condition should decrease with age. As the lower panel of Figure 29 shows, the data confirm this prediction as well.

This same set of hypotheses and assumptions yields parallel predictions as to the relative efficacy of verbal (sentence) and visual (action pictures) forms of elaboration as a function of age. Data relevant to these predictions have been reviewed elsewhere (Rohwer, 1970); in brief, the evidence presently available appears to be in accord with the predictions. That is to say, the younger the child, the more effective sentence elaboration is relative to action-picture elaboration, and, with increasing age, the less do sentence descriptions of action pictures improve performance over that produced by action pictures alone. Other alternative explanations of these data have been proposed (Paivio, 1970; Palermo, 1970; Reese, 1970) but a choice among the alternatives must await further experimentation.

It would be both useful and satisfying to be able to offer for consideration at this point one or two parsimonious theoretical generalizations that would simultaneously summarize all of the foregoing research and suggest new empirical implications. Unfortunately, this objective is still beyond reach. Nevertheless, several assertions can now be made about the role of mental elaboration in learning and it is possible to single out a few issues that are particularly in need of resolution.

First, consider a summary account of the assertions. Mental elaboration, that is, imaginative conceptual activity, has a demonstrably powerful effect on learning efficiency. The kind of experimental analysis permitted by the method of manipulating external analogues of hypothetical elaborative activities demonstrates that effective elaboration has extraordinarily specific properties. In the case of verbal elaboration, these properties are both syntactic and semantic in character while in the case of visual elaboration they seem to be both spatial-relational and episodic-thematic. The parallels between the effective properties of the two modes of elaboration are striking and these very parallels suggest that either the one mode accounts for the other or that some third kind of underlying process accounts for both. It may be speculated that such an underlying process is one that relates semantic language features with visual imagery, thus yielding the properties of memorable storage units. Research has also shown that elaborative phenomena are sensitive to certain task variables such as stimulus mode, response mode, type of test cue and pacing.

The major issues still in need of attention include the theoretical problem of finding a unifying account of the effects of visual and verbal elaboration as well as that of clarifying the matter of visual and verbal dominance relations when viewed developmentally. Finally, it is of interest to consider the question whether or not the effective forms of elaboration thus far identified serve to advance attempts to account for a variety of individual differences in learning proficiency.

## Mental Elaboration and Learning Proficiency

The shift at this point from the terms learning efficiency to the terms learning proficiency signals a marked shift in emphasis. Thus far, our primary concern has been to identify types of elaboration that are generally effective and to specify the properties responsible for their effectiveness. The shift is to a concern with the question whether or not the degree of efficiency that characterizes the learning of individuals varies in systematic ways as a function of elaboration variables. Thus, proficiency refers to enduring patterns of learning efficiency in individuals and in groups of similar individuals.

Recall that one of the starting points for the entire line of research reported here was the suspicion that individuals differ in their use of imaginative conceptual activity as a means of acquiring information. If this suspicion is warranted, then it should be possible to account for certain kinds of individual differences in learning proficiency in terms of corresponding differences in elaborative activities. Although there are numerous characteristics of individuals that warrant this kind of analysis, we will confine our attention to only three of these: age, IQ and a combination of ethnicity and socioeconomic status (SES).

### Age, Elaboration and Learning Proficiency

Considerable discussion has already been devoted to this topic in connection with the problem of developmental differences in the dominance of verbal and visual modes of elaboration. Even at the risk of some redundancy, however, some additional comment is appropriate. The initiating phenomenon is the observation that learning proficiency improves with age, at least insofar as it is indexed by performance on PA tasks (Jensen & Rohwer, 1965). College students learn more efficiently than twelve-year-olds and twelve-year-olds learn more efficiently than six-year-olds. The assumptions made here about activities should account for a portion of the age-related variance in learning proficiency.

Thus far, the amount of empirical information available for evaluating this hypothesis is sparse indeed. But it is possible to identify two age contrasts that display differences in learning proficiency and concurrent differences in elaboration. The first of these, of course, is the contrast between four- and eight-year-old children with respect to both learning efficiency and elaboration. Here are the facts: eight-year-olds are more proficient learners than four-year-olds; when elaboration is provided, both visual and verbal forms facilitate learning in eight-year-olds but only verbal forms facilitate learning in four-year-olds; the combination of visual and verbal forms of elaboration adds a detectable increment to the performance produced by the verbal form alone in four-year-olds, but adds nothing for eight-year-olds (Rohwer, 1970). In addition to raising the issue of developmental dominance of the verbal mode of elaboration, these facts suggest that over the age interval four to eight, children come more and more to generate their own forms of mental elaboration in response to presented materials.

The second age contrast is that between elementary school children and college students. Since college students are generally more proficient at learning PA lists than school children, the expectation is that they also engage in more autonomous elaborative activities than school children. If so, it follows that the presentation of learning materials in elaborated forms should produce less facilitation in the older than in the younger subjects, as compared with performance in a control condition where the task is presented by an ordinary PA procedure. In two studies of sentence elaboration where direct comparisons are possible between sixth-grade children and college students (Suzuki & Rohwer, 1969; Suzuki, 1969) the results conform to this expectation; the sentence conditions did facilitate learning for children but not for adults. Similarly, Bower (1969) has reported that sentence elaboration provided by the experimenter does not facilitate learning in college students relative to an ordinary control condition. This experiment, however, included a second control condition in which subjects were instructed to rehearse each of the noun pairs as it was presented, thus effectively filling all presentation intervals with rote activity designed to prevent autonomous elaborative activity. Performance in the rehearsal control was significantly inferior to that in the ordinary control and to that in the presented sentence condition, confirming the presumption of spontaneous elaboration in college students.

In contrast to the case of presented elaboration, experiments in which elaboration instructions have been manipulated yield significant facilitation even for college students (Bower, 1969; Paivio & Yuille, 1967). Facilitation attributable to sentence instructions has also been reported for children (Jensen & Rohwer, 1965; Milgram, 1967a). Even though a direct age comparison has not yet been made in a single study, the magnitude of the instructional effect seems to be smaller in college students than in school children.

Thus, the assumption that more proficient learners (college students) are characterized by more autonomous elaborative activity than less proficient learners (school children) is consistent with relevant data presently available. Furthermore, Martin (1967), using a post-learning interview method, found that the frequency of reported elaborative activity increased significantly with grade level, across samples of fourth-, sixth- and eighth-grade children. Even though the evidential case is not yet entirely compelling, it does appear that age-related increases in learning efficiency are attributable to concurrent changes in elaborative activities.

IQ, Elaboration and Learning Proficiency

Although modest in magnitude, reported correlations between IQ and performance on PA learning tasks indicate that a positive relationship obtains between IQ and learning proficiency. In one study, for example, within-sample correlations between Peabody Picture Vocabulary Test (PPVT) IQ and PA performance for kindergarten, first-, and third-grade children averaged .31 (Rohwer, in press). In a similar study with a different age range (three and a half to five and a half year olds), the average correlation between PPVT IQ and PA performance was .34 (Rohwer, 1967).

In another study, the PA performance of institutionalized retardates was compared with that of kindergarten, first-, third-, and sixth-grade children sampled in equal numbers from schools serving high-SES white populations and schools serving low-SES Negro populations (Rohwer & Lynch, 1968). Within each sample, the 24-item PA list was administered to independent groups under each of four conditions: Names-Still, Names-Action, Sentence-Still and Sentence-Action. In all samples, the three elaboration conditions produced better performance than did the Names-Still condition and the patterns of facilitation observed were virtually the same for the retardates as for the school children. The overall level of performance in the retardate sample, however, was inferior to that in every other sample, including that of the low-SES kindergarten children whose mean mental age (MA) was substantially below that of the retardates.

One comparison of particular interest in the study was that between the high-SES, third-grade children and the retardates, with whom they were matched for MA. The performance of the retardates was significantly inferior to that of the third-graders but the pattern of differences produced by the various conditions, elaboration and control, was highly similar. We interpreted this outcome as providing support for Zigler's (1967) contention that normals and retardates of equal developmental level (MA) are characterized by comparable cognitive structures. In addition, however, we interpreted the inferior level of absolute performance in the retarded sample as a contradiction of Zigler's inference that equivalent cognitive structures imply equivalence of learning efficiency. It seems patent to me that equivalence of learning rate is not a necessary consequence of structural equivalence but Zigler (1969) has taken sharp issue with this interpretation.

One other feature of the results of this experiment deserves mention at this point even though it will be treated again in the following s tion, namely, the fact that no significant differences were observed between the performance of the high-SES white and the low-SES Negro children. This result was surprising in view of the fact that the average IQ of the high-SES white samples was substantially above that of the low-SES Negro samples where the mean was so low as to imply that several children had been sampled from the retarded range. This result, namely, equivalence of learning efficiency between the two SES groups, suggests that cultural and familial retardation can be separated in terms of performance on learning tasks; familial retardates would be expected to be less efficient learners than cultural retardates of the same MA (cf. Rapier, 1968).

Even though the assertion has not been completely established that some of the individual differences variance shared between IQ and PA learning can be accounted for in terms of individual differences in elaborative activities, the case appears to be a relatively strong one.

## Ethnicity, SES, Elaboration and Learning Proficiency

In comparison with Age and IQ, it is a severely complicated problem indeed to relate individual differences in ethnicity, SES and learning proficiency to comparable differences in elaborative activities. The major source of difficulty is created by the fact that various ethnic and SES populations have been shown to be equivalent in learning proficiency as measured by performance on PA tasks (Semler & Iscoe, 1963; Rohwer, Lynch, Levin & Suzuki, 1968; Green, 1969). The problem is made even more severe for an elaboration theory of individual differences by the fact that equivalence of performance among such populations is more often observed when the PAs are administered without elaboration than when the elaboration is provided.

In contrast to the results obtained when PA learning serves as the index of learning proficiency, performance on school achievement tests, often presumed to be measures of long-term learning proficiency, is strongly associated with ethnic and SES differences; this association is comparable in strength to that usually obtained between IQ and ethnicity-SES. Consider an example. Green (1969) recently conducted a study of fourth-grade Negro children in which equal numbers of subjects were sampled from low- and middle-SES populations. The average total reading score of the middle-SES sample on the Stanford Achievement Test was 72.8 as compared with an average of only 46.3 for the low-SES sample. Similarly, the average IQ (Lorge-Thorndike) of the middle-SES group was 96.1 while that of the low-SES sample was 79.1.

Given these data, it might be argued that SES-related differences in school learning are accounted for by comparable SES-related differences in IQ, especially if it is granted that IQ is a measure of learning proficiency. Before this assumption is granted, however, it deserves closer examination, principally because of the fact that neither IQ tests nor school achievement tests require the student to engage in learning. To the contrary, both kinds of tests ask the student to recall and apply information he has acquired prior to the testing session itself. Thus, the question is whether or not differences in learning proficiency when this trait is directly measured by tasks that require new learning.

The Green (1969) study itself provides contradictory answers to this question. In addition to the scores already reported for the middle- and low-SES samples, all of the children were administered three other tests: Raven's Progressive Matrices (Raven); a digit-span task; and a PA task. The digit-span task required the child to listen to random strings of digits, varying in length from three to nine numerals, and to repeat them immediately. The PA task was a 20-item list of paired objects presented on movie film under a condition comparable to that referred to previously as Name-Still. The average performance of the middle-SES sample on both the digit-span task and the Raven was markedly better than that of the low-SES sample. If either of these tests qualifies as a direct measure of learning proficiency, then it seems warranted to conclude that IQ also measures learning proficiency and that IQ differences account for SES-related differences in school achievement. The results for the PA test, however, are in direct opposition to this conclusion. The mean number of correct responses for the middle-SES sample was 24.8, while for the low-SES sample the mean was 24.1. Thus, if the PA task is construed as a direct measure of learning proficiency, IQ differences cannot be said to account for SES-related differences in school learning.

A number of other studies have produced results that are equally perplexing. In general, when the samples selected are six years of age or older, differences between SES and ethnic populations are not detected on tests of PA learning even though the populations may be radically different in terms of performance on school achievement and IQ tests. When independent-groups designs are used, the equivalence of high-SES white and low-SES Negro samples holds for elaborated as well as non-elaborated conditions of PA learning (Rohwer, Lynch, Levin & Suzuki, 1968; Semler & Iscoe, 1963). The task of free recall learning has also been administered to samples of high-SES white and low-SES Negro children (Glasman, 1968; Jensen & Frederiksen, this report). Two kinds of item lists have been used in these studies, that is, categorized and uncategorized lists, and the results obtained depend entirely on which kind of list is administered. That is to say, marked SES differences emerge in performance on categorized lists whereas the SES samples perform at equivalent levels on uncategorized lists.

Jensen (1969a) has proposed a model to account for the discrepancies in results among the various studies of SES-related differences in learning proficiency. The model posits two distinguishable varieties of learning ability: associative and conceptual. Associative learning is characterized as involving " . . . the neural registration and consolidation of stimulus inputs and the formation of associations. There is relatively little transformation of the input, so there is a high correspondence between the forms of the stimulus input and the form of the response output." (Jensen, 1969, pp. 110-111). Tasks such as digit span, serial learning, free recall of uncategorized lists and PA learning are thought to measure associative learning ability. Conceptual learning ability, in contrast, is held to involve considerable transformation of stimulus input and is measured by performance on tasks such as that of the Raven.

Jensen suggests that a review of available empirical evidence demonstrates that high- and low-SES groups differ in performance on conceptual learning tasks but not on associative learning tasks. Furthermore, he notes that in some studies, the correlation between performance on the two varieties of tasks is very low for low-SES samples and moderately high for high-SES samples. From these facts, he hypothesizes that associative learning ability is distributed equally among the various SES populations but that conceptual learning ability is not. If the hypothesis is true, it is reasonable to recommend, as Jensen does, that school subjects should be taught to low-SES children in a form suitable for acquisition by means of associative learning and to high-SES children in a form amenable to conceptual learning processes.

Both the model proposed by Jensen (1969a) and its implications are reasonable and important for psychology as well as for education. But it has one major flaw at its source, namely, that it does not fit the data. For example, the model identified digit-span tasks as measures of associative learning ability and yet such tasks reveal striking differences between SES samples (Green, 1969). There are large differences between SES groups in performance on a test like the PPVT which simply requires the recall of verbal labels for pictured objects, hardly a highly conceptual transformational process. In addition, some available data to be reported shortly disconfirm the notion that associative and conceptual abilities are more highly related in high-SES than in low-SES samples.

Furthermore, the model is difficult to support with respect to its identification of tasks such as PA learning and the free recall of uncategorized lists as measures of associative, not conceptual learning ability. Indeed, one of the principal theses of the present paper is that conceptual activity is centrally involved in determining PA learning proficiency and the evidence to support this contention is substantial. With regard to free recall tasks, there is also considerable evidence to the effect that they provoke conceptual activity whether the lists are composed of categorized or uncategorized items -- otherwise, it is extremely difficult to account for the phenomena of clustering and subjective organization in responses to uncategorized lists.

In view of these difficulties in the model proposed by Jensen (1969a), I have proposed an alternative one (Rohwer, 1969) which specifies a two-dimensional space within which various intellectual tasks can be located. A schematic display of this model is presented in Figure 30. The poles of one dimension designate the kind of

Insert Figure 30 about here

conceptual activity likely to be elicited by a task: formal or

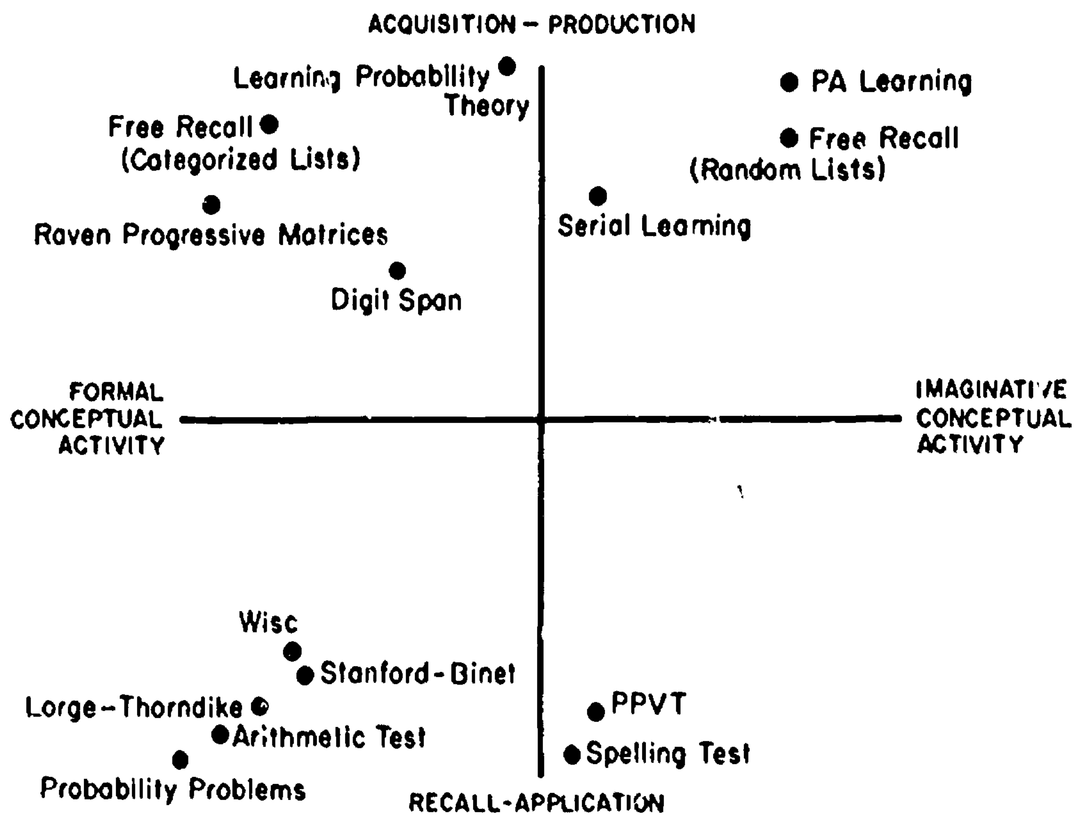



Figure 30. A two-dimensional model for classifying various intellectual tasks.

imaginative. This distinction corresponds to that made in introducing the topic of elaboration in the present paper. Successful performance on tasks that tend to provoke formal conceptual activity requires the acquisition and application of a set of relatively explicit rules capable of exhaustively describing either the materials to be processed or the operations necessary for the completion of such processing or both. Incidentally, it is an almost unfailing characteristic of such rule systems that they are legitimized by cultural consensus. Imaginative conceptual activity, in contrast, is often quite idiosyncratic in character, involving the invention of ad hoc ways of processing and transforming information. These are the kinds of processes that are heavily involved in elaborative activities that successfully facilitate PA, free recall, and even serial learning (Bower, 1969; Levin & Rohwer, 1968). The assumption which generates this dimension is that proficient learners engage in conceptual activity in performing any task that demands the acquisition or production of new information -- proficient learners are not rote learners.

The second dimension refers to the type of behavior demanded by the task, ranging from acquisition-production behaviors at the one pole to recall-application behaviors at the other pole. This dimension is, of course, a crucial one for rationalizing the results of SES-related differences in task performance -- if the information or the skills demanded by a recall-application task have not been learned previously, effective performance on the task is clearly impossible.

It is not unreasonable to suppose that there are reliable individual differences with respect to both of these dimensions. Some persons are probably predisposed toward formal conceptual activity while others are predisposed toward imaginative conceptual activity; some are probably better at acquisition-production while others are better at recall-application. Furthermore, individual differences such as these might well be expected to be quite pronounced within definable populations, that is, within SES and ethnic groups. It is equally reasonable to suppose, however, that there may be differences between groups with respect to their propensity for one or the other of the two kinds of conceptual activity. The results reported by Stodolsky and Lesser (1967) point in this direction.

An inspection of Figure 30 reveals that the placement of tasks in this model differs in some important respects from the placement of the same tasks in the model proposed by Jensen (1969a). Here, differences between high-SES white samples and low-SES Negro samples have been reported for all tasks save those located in the imaginative-acquisition quadrant. Thus, the model provides, at a minimum, a partitioning of tasks that conforms with available empirical evidence and with the theoretical interpretations outlined herein of the processes underyling performance on these tasks.

The present model can generate a number of interesting predictions about the interaction of SES, ethnicity, task requirement (recall vs. acquisition) and conceptual activity (formal vs. imaginative). One of these predictions, of course, is that low-SES Negro populations will differ from high-SES white populations on recall-application tasks and on tasks that require formal conceptual activity but not on acquisition tasks that require imaginative conceptual activity. The fact that the hypothesis fits the data presented thus far is not impressive since it was constructed precisely to do this. An evaluation of its adequacy awaits the conduct of new empirical tests.

The model also has pronounced educational implications, however, and these are worth brief mention at this point. It implies that learning, of whatever variety, proceeds best when conditions of learning are sufficient to elicit conceptual activity in the learner, whether the kind of activity called for is formal or imaginative. It does not imply that some subject matters should be taught to some students by engaging them in rote activity and to other students by engaging them in conceptual activity. Instead it implies that for some students a particular subject matter should be presented for learning in such a way as to permit acquisition by means of imaginative conceptual activity while for other students the subject matter should be presented so that it can be acquired by means of formal conceptual activity. The model also implies that for low-SES students care should be taken to insure that ample opportunities are provided for acquiring information and skills missed because of inadequate early environmental experience, and, of equal importance, these opportunities should be tailored to the students' relative propensities for formal or imaginative conceptual activity. Simply, the argument is that a given subject matter can be mastered efficiently either by the route of formal or by the route of imaginative conceptual activity, depending on the propensities of the students being taught; the corollary argument is that the achievement of mastery by means of rote activity is probably inappropriate for all students.

The remaining two papers in this report describe studies that are relevant to both the model proposed here and to a number of the issues it raises.

# Ethnicity-SES and Learning Proficiency[1]

William D. Rohwer, Jr., Mary Sue Ammon,

Nancy Suzuki and Joel R. Levin

Currently, one of the most visible of educational phenomena is the marked discrepancy in school achievement between Black children from families of low socioeconomic status (SES) and White children from high-SES families. It is commonly reported that differences between these two populations in performance on standardized achievement tests are as large as forty to fifty points on percentile scales (Coleman, 1966; Rohwer, 1969; Wilson, 1963). The question is, what accounts for such observed differences in achievement?

One straightforward answer is that the two populations, high-SES White and low-SES Black, differ in learning proficiency. Empirical support for this answer would consist of evidence showing similar differences between the two populations on measures of learning proficiency that are at least operationally independent of school achievement tests. If it is assumed that intelligence tests index learning proficiency in a relatively unbiased manner, then such evidence is readily and plentifully available. Differences in IQ between high-SES White and low-SES Black children repeatedly have been shown to be in the same direction and of approximately the same magnitude as differences in standardized achievement test scores (Nichols, 1969). Accordingly, it may be concluded that differences in learning proficiency explain observed differences in school achievement between the populations.

The problem with this explanation is the assumption that IQ indexes learning proficiency. Intelligence tests rarely require the child to engage in learning; they require him to give evidence that he has learned previously. Thus, rather than commanding immediate acceptance, the assumption needs empirical support of the kind that would be provided by a demonstration that the scores yielded by intelligence tests parallel scores yielded by tasks that directly involve the child in learning.

The number of relevant studies presently available is very small. Few investigations have been undertaken in which learning tasks and intelligence tests have been administered to samples drawn from both high-SES White and low-SES Black populations of school children. Those which have been conducted, however, cast considerable doubt on the validity of the assumption that IQ is an unbiased measure of learning proficiency. Semler and Iscoe (1963) observed the performance of White and Black elementary school children on paired-associate (PA) learning tasks and on the Wechsler Intelligence Scale for Children (WISC). They found substantial race differences in WISC IQ but not in paired-associate learning efficiency. Another PA learning experiment (Rohwer, Lynch, Levin & Suzuki, 1968) failed to detect differences between high-SES White and low-SES Black children in any one of six variations of presentation method. Similarly, in a study confined entirely to Black elementary school children, Green (1969) has reported finding no significant

[1]We wish to acknowledge the indispensible assistance of the teachers and children in those schools where the data were collected as well as the contributions of Miss Helgola Ross and Miss Marsha Lynn in administering the various tests and of Dr. Douglas Penfield in analyzing the results.

SES differences in PA learning efficiency but marked SES differences in Lorge-Thorndike IQ. Jensen (1968b), using a series of digit-span tasks along with the children's form of the Raven Progressive Matrices test, detected large differences between high-SES White and low-SES Black children on the Raven but not on digit span. Thus, these direct measures of learning proficiency commonly fail to reveal population differences of precisely the kind that would be expected on the assumption that IQ is a valid and unbiased index of learning proficiency.

Nevertheless, the issue is not as simple as this brief review indicates. The relative performance of high-SES White and low-SES Black children on intellectual tasks fluctuates as a function of a number of specific variables; chief among these are task differences and the chronological age of the Ss. When paired associates is the method used, the typical results are that significant amounts of between-groups variance are regularly associated with population membership among three-, four-, and five-year-old children, occasionally in six year olds, and rarely in children seven years or older (Rohwer, 1967, Experiments XII and XIII; Rohwer & Lynch, 1968; Rohwer, Lynch, Levin, & Suzuki, 1968; Semler & Iscoe, 1963). The comparable developmental function has a very different form when digit-span tasks are used. In preschool children, that is, in three, four, and five year olds, the performance of high-SES White and low-SES Black children is virtually equivalent (Jensen, 1968b) whereas in fourth-, fifth-, and sixth-grade children, digit memory among high-SES Whites is markedly better than among low-SES Blacks. Furthermore, digit-span performance is considerably better among high- than among low-SES Black children at the third-grade level (Green, 1969). The task of free-recall learning also yields results showing a divergence between population groups with increasing age. Glasman (1968) presented categorized lists of familiar objects to high- and low-SES kindergarten and fifth-grade children with the result that recall was equivalent for the kindergarten children but not for the fifth-grade samples where the high-SES groups scored substantially higher than the low-SES groups.

In view of both the issue at stake and the evidence presently available, several questions warrant clear answers.

Paired-Associate Test Reliability. One of these is raised by the use of a learning task, in this case a PA task, to estimate individual and group differences in learning efficiency. When a task is to be used for this purpose, it is important to know its reliability, but, in contrast to intelligence tests, such information is rarely available for learning tasks. Accordingly, the present experiment was designed to yield estimates of the reliability of the PA task included in the test battery.

Populations Differences and Varieties of Learning. Another question pertains to an hypothesis proposed by Jensen (1969a) regarding a difference between high-SES and low-SES children in the organization of learning abilities. Jensen distinguishes two broad varieties of learning ability, associative (Level I) and conceptual (Level II). Presumably, Level I abilities are principally exercised on tasks that require the verbatim reproduction of the information originally presented for learning, tasks such as digit-span and PA learning. In contrast, Level II abilities are elicited by tasks that require the S to transform the information given in order to produce responses that are counted as being correct. Jensen

(1969a)has identified the Raven Progressive Matrices as an exemplar of a task requiring Level II ability for successful performance. On the assumption that Level I abilities are distributed equally among the two populations, high-SES White and low-SES Black children, whereas Level II abilities are not, two interesting predictions follow. The first is that populations differences should be detected on tasks that principally elicit Level II abilities but not on tasks that elicit Level I abilities (e.g., Raven vs. PA tests). However, the foregoing review suggests that any assertions about how populations differ as a function of tasks must be qualified in terms of the ages of the Ss sampled. In the present experiment, this was accomplished by administering three different kinds of tasks to both high-SES White and low-SES Black children drawn from three age levels. One kind of task, the PA test, has previously revealed populations differences only for young children (Rohwer, 1967, Experiment XIII); another was selected as an exemplar of Level II tasks (Raven Progressive Matrices); and a third, Peabody Picture Vocabulary Test (PPVT), was selected as representative of widely used, relatively brief, IQ tests. The second prediction derived from the Jensen model is that the magnitude of the correlation between performance on Level I and Level II tasks should be greater for high-SES White than for low-SES Black children. By design, the present study provides an empirical test of both these predictions.

The apparently singular issue whether or not there are populations differences in PA learning proficiency may be formulated in several different ways. The simplest of these has already been considered, namely, the question, are there populations differences in the efficiency of learning lists of paired associates? Another formulation of the issue concerns the question whether or not there are populations differences in the amount of profit derived from the experience of performing on paired-associate tasks prior to the learning of some subsequent list, that is, are there populations differences in the efficiency of nonspecific transfer or learning to learn (LTL)? Still another formulation concerns the possibility that the magnitude of populations differences in learning efficiency varies as a function of the manner in which the learning materials are presented. And, the final formulation frames the issue in terms of the efficiency of recall rather than in terms of the efficiency of original learning.

Learning to Learn. Do high-SES White and low-SES Black children differ in the amount of transfer, namely, learning to learn (LTL), that accrues from performance on successive PA lists? It might be argued that even if the two populations do not differ in single-list learning efficiency, they do differ in a capacity more vital for successful school learning, the capacity to transfer what has been learned from one instructional sequence to performance in another similar sequence. To assess this possibility, Ss in the present study learned four different PA lists.

Methods of Presentation. Are the results bearing on the issue of populations differences specific to a particular method of presenting the pairs? In order to provide at least a limited answer to this question, the PA test was constructed of lists of noun pairs within which the PAs were presented in one or another of five different ways. Three of these item types were selected because of their demonstrable effect on learning efficiency. It has been shown that noun pairs depicted in the form of the objects to which they refer are learned more easily when they are (a) presented in the context of sentences, or (b) presented in the form of action episodes relating the two members of each pair, than when the objects are simply shown as still pictures and named aloud for the S (Rohwer, Lynch, Levin & Suzuki, 1967). The facilitating effect of action pictures and sentence verbalization has been shown to hold for low-SES Black children as well as for high-SES White children at all grade levels assessed, kindergarten, first, third and sixth (Rohwer, Lynch, Levin & Suzuki, 1968).

Rohwer (in press) has treated these methods of presentation as external analogues of hypothetical internal mental activities engaged in by persons who are efficient learners. The notion is advanced that successful PA learning is promoted by the elaboration of the raw elements to be acquired so as to invest them with membership in a single semantic set, either by lodging them in the same linguistic unit, as in a sentence, or in the same pictorial unit, as in an action episode. Independent evidence in support of this notion is provided by experiments in which similar facilitation effects have been associated with instructions to elaborate noun pairs (Bower, 1968, Jensen & Rohwer, 1965) and with S reports of spontaneous elaboration at the completion of PA learning (Bugelski, 1962; Martin, 1967; Runquist & Farley, 1964).

In addition to those forms of elaboration that can be construed as serving to form semantic sets, Rohwer (1968) has also described two other forms that are more elementary in nature than the use of sentences or action imagery, but which are parallel in that one is verbal in character and the other is pictorial. The first of these primitive forms is that of generating a verbal label or name for pictorial stimuli and the second is that of generating a pictorial image of the referents of auditory stimuli.

The assessment of the effects of each of these four forms of elaboration on PA learning in children requires the use of five different ways of presenting noun pairs: names of objects, pictures of objects, pictures of named objects, pictures of objects along with sentences containing the object names, and action pictures of named objects. The importance of each of the forms of elaboration for efficient learning can then be determined by comparing every one of the remaining four item types with that consisting of pictures of named objects. Accordingly, the PA lists used in the present study included pairs representing all five item types.

The reason for manipulating the variable of PA item types is that it permits a specification of the conditions under which populations differences in learning efficiency occur, if they occur at all. Furthermore, it has been hypothesized (Rohwer, 1968) that if low-SES Black children have any deficiency in learning skills, it is a relatively weak propensity to elaborate the materials to be learned. From this hypothesis, the prediction follows that populations differences are more likely to be detected on the less elaborated item types and not on the ones where the elaboration is furnished in the learning materials themselves.

Retention. The final question is concerned with the possibility of populations differences in the efficiency with which information already learned can be recalled after a lapse of some specified amount of time. It might be reasoned that the inferior performance of low-SES Black children on school achievement tests is due to limited capacity for initial learning, to limited capacity for retention, or to both. Thus provision was made for assessing the number of PAs retained as well as the number initially learned.

## Method

### Subjects

The total sample numbered 288 children drawn in equal numbers from the six populations defined by the classification factors of Grades (K, 1, 3) and SES-Ethnicity (high-SES White, low-SES Black). The populations were defined by the manner in which they were located as follows. The study was conducted in two communities known for the ethnic homogeneity of their school populations, one White, the other Black. Within the community from which the Black sample was to be selected, a particular school was chosen in accord with the rule that it served a set of census tracts in which the households were clearly classifiable in terms of SES. The variables available in census information were: median income, median education level, percentage homeowners, average value of homeowners' dwellings, average rent of other dwellings, ratio of "deteriorating" and dilapidated houses to "sound" houses, and a crowding index. After the schools were designated, sampling within grade levels was conducted by randomly selecting 24 males and 24 females from a list of all children enrolled in that grade. Within the groups originally selected, children absent on scheduled testing days were replaced from a list of randomly chosen alternates. There was a total of 7 such cases among the high-SES White samples and 19 among the low-SES Black samples. Chronological age information for each of the samples is given in Table 31.

- - - - - - - - - - - - - -
Insert Table 31 about here
- - - - - - - - - - - - - -

### Tasks and Materials

Three different kinds of tasks were administered to every child individually: Peabody Picture Vocabulary Test (PPVT), Form B; Coloured Raven Progressive Matrices (CPM); and, four paired-associate (PA) lists.

PPVT. The test consists of a booklet with an array of four pictures appearing on every page. As a page is exposed to the view of the S, E utters a word for which one of the four pictures is a referent. The S's task is to point to the picture depicting the referent. The procedures as described in the manual (Dunn, 1965) were followed for both the administration and the scoring of the test. Thus, three measures were obtained for each child: raw score, mental age (MA) and IQ. Based on the performance of the standardization sample, the alternate forms reliability of the test for children in grades K, 1 and 3, respectively are: .73, .69, .79.

CPM. The book form of the CPM was used, and, with a few exceptions, the procedures described in the manual (Raven, 1960) were followed in administering the test: the wording of the instructions was modified slightly to conform to American usage; and, prior to introducing the child to the first test problem in the book, he was given four practice problems in a board format. The board problems were used to present the instructions

Table 31

Chronological-age Means and Standard Deviations (in months)

as a Function of Grades and Populations

| Grades | Populations | | | |
|---|---|---|---|---|
| | High-SES White | | Low-SES Black | |
| | $\bar{x}$ | s | $\bar{x}$ | s |
| K | 68.02 | 5.01 | 70.42 | 4.32 |
| 1 | 77.94 | 3.76 | 81.58 | 4.69 |
| 3 | 102.62 | 4.32 | 106.25 | 4.74 |

for the test itself with the intention of making clear to the child that his choice of a figural alternative was for the purpose of completing the given pattern. The CPM consists of a total of 36 problems divided into three sets (A, AB, B) of twelve problems each. Although the manual does not provide information sufficient to lead to a confident estimate of the reliability of the test on populations such as those sampled here, Raven (1960, p. 15) notes that the test-retest coefficient is approximately .65 for children of the age range recruited for the present study.

PA test. Each of the four PA lists is comprised of 25 noun pairs administered in accord with a study-test method for a total of two complete trials, that is, two study and two test trials. Within every list, the study-trial materials consist of the various pairs presented in one or another of five different ways, so that each of the five Item Types are represented by five pairs in each list. Since the types of items are distinguished from one another in terms of both auditory and visual features, they are presented by means of videotape to hold the test constant across administrations. The five Item Types are: Nouns, in which each noun pair is presented aurally; Pictures, in which each noun pair is represented by a picture of two objects; Nouns-Pictures, where pictures of object pairs are presented with the aural presentation of their labels; Sentences-Pictures, consisting of pictures of objects whose names are presented in a sentence describing some kind of interaction between them; and, Nouns-Action, where the visual signal literally depicts an interaction between the two objects shown while the names of the objects are presented aurally.

In order to permit an unequivocal attribution of expected differences among Item Types to corresponding differences among the presentation methods, a pretest was conducted to estimate the difficulty of each of the 100 pairs in the four lists. All 100 pairs were prepared for presentation by the Nouns-Pictures method and were randomly assigned to four lists of 25 pairs each. The four lists were administered to samples drawn from populations similar to those sampled for the present study itself. The average number of correct responses given for each pair was used to estimate pair difficulty and the 100 items were ranked accordingly. This ranking was divided into twenty levels of five items each and one item from each level was assigned to each of the Item Types. Finally, the pool of twenty pairs for each Item Type was randomly subdivided into groups of five items and assigned randomly to the four lists that constituted the PA test. The order of the pairs on the tape is random with respect to Item Type with the restriction that all types are represented once in each sequence of five pairs. During the study trials, successive pairs occur at a 4-second rate.

The test trial materials for each list are also recorded on videotape. For every noun pair, either an object or a noun or both are presented. These stimuli appear at a 4-second rate, but in an order different from that of the study trials. As in the case of the study-trial materials, each Item Type is represented by a test stimulus in every sequence of five stimuli.

The instructions for the PA task informed Ss about the various Item Types and urged them to learn each pair in such a way that they could supply the missing pair member on test trials. To clarify the instructions, a five-item practice list, with one pair representing each of the five Item Types, was presented prior to each of the sets of two 25-item lists. The practice list was administered repeatedly until S attained a criterion of at least three correct responses.

## Procedure

All Ss received the CPM, the PPVT, and the PA test during three separate testing sessions. The first session was devoted to the CPM, the second to the PPVT and two of the PA lists, and the third to the remaining two PA lists. The first two sessions were separated by an interval of varying length, from two to five days, but in every case sessions two and three were separated by a two day interval. The constancy of this latter interval is important because the third session always included the administration of the test trial materials from each of the two PA lists learned in the second session. The purpose of this procedure was to assess PA retention as a function of the various classification variables and of Item Types. Following the administration of the two new PA lists, the third session concluded with the presentation of one test trial for each of the two PA lists learned during the previous session.

## Design

The analysis of variance design common to all three tasks was a three-way factorial, Grades (K, 1, 3), Populations (high-SES White, low-SES Black), and Sex (males, females). In the case of the PA tasks, this basic design was augmented to permit the assessment of a number of sources of within-subjects variance. In designating these sources, it is necessary to distinguish between the dependent variables of original learning and recall. With respect to original learning, the additional variables were: Item Types (Nouns, Pictures, Nouns-Pictures, Sentences-Pictures, Nouns-Action); Trials (1,2); and, Practice (first, second, third and fourth lists). The variable of Practice allows for an assessment of amounts of generalized transfer as a function of the subject classification variables of populations and grade level. It was possible to assess the effects of Practice free of the influence of differences in difficulty among the four lists because the order in which the lists were administered was completely counterbalanced within each of the six samples. That is to say, two Ss in each sample were randomly assigned to each of the 24 possible list orders.

In addition to the status variables of Grades, Populations and Sex, the design for the analysis of the recall data included Item Types and Lists (1,2). The recall trials for each of the two lists were always administered in the order that the lists were presented during original learning. Once again, because of counterbalancing with respect to list order in original learning, all lists were equally represented in the first and second recall positions. It should be noted that neither the recall of the first nor of the second list can be construed as providing a measure of simple retention. Since the learning of the second, third, and fourth lists intervened between first-list learning and first-list recall, retention of the first list was subject to retroactive interference effects. Similarly, since the learning of the first-list preceded the learning of the second, second-list recall was subject to proactive interference effects as well as to possible retroactive effects from the learning of the third and fourth lists. Furthermore, second-list recall was also subject to possible interfering effects from the activity involved in the immediately preceding attempt to recall the first list. Although the variety of possible interfering and facilitating effects complicates any interpretation of relative retention for first and second lists, between samples contrasts for both measures are meaningful in view of the fact that all Ss were subject to the same effects.

The designs for the correlational analyses were straightforward. All, including both the reliability studies of the PA test and the intertask correlations, were performed within the samples yielded by the combination of the factors Grades and Populations. The variables entered into these analyses were: PPVT raw score, CPM raw score, PA total score, and total scores for each of the five PA Item Types.

## Results

Paired-Associate Test Reliability. The first aspect of the results to be examined is concerned with the reliability of the PA test. The method of alternate forms was used to produce the reliability coefficients. For each of the six samples, six such coefficients were calculated, one for each of the Item Types and one for performance summed across Item Types. In every case, the scores consisted of the numbers of correct responses given on the test trials summed across two of the four lists. For all Ss, one form of the test was defined as the first two lists administered and the other form consisted of the remaining two lists. By this procedure, list differences were balanced across Ss. Thus, the maximum total score on either form of the test was 100 and the maximum score for each Item Type was 20.

The results are presented in Table 32 as a function of Grades and Populations. The reliability coefficients for the total score on the PA

- - - - - - - - - - - - - - -

Insert Table 32 about here

- - - - - - - - - - - - - - -

test are acceptably high for most of the samples. In all cases, it may be that the coefficients reported underestimate the maximum reliability of some particular pair of alternate forms available among the four PA lists used. All possible pairings of the four lists are equally represented in the coefficients that have been calculated so that the only feature of the two forms common to all Ss is that one form consists of the first two lists administered while the other form consists of the second two. Even so, the reliability of the total score is quite comparable with the reliabilities reported for the PPVT and the CPM.

As would be expected, the reliability coefficients produced by the individual Item Types are generally lower than those for total scores. When any one of the Item Types is treated as a test in itself, the factor of test length becomes important; each such test is only ten items long, even when two of the full PA lists are involved.

Population Differences and Varieties of Learning. Having established the relative comparability of the three tasks, PA, PPVT, and CPM, with

Table 32

Reliability Coefficients for the PA Test as a Function of Grades, Populations, and Item Types

| Samples | Nouns | Pictures | Nouns-Pictures | Sentences-Pictures | Nouns-Action | Total |
|---|---|---|---|---|---|---|
| High-White K | .45 | .24 | .28 | .28 | .32 | .54 |
| Low-Black K | .59 | .56 | .57 | .67 | .76 | .87 |
| High-White 1 | .45 | .56 | .54 | .42 | .63 | .80 |
| Low-Black 1 | .38 | .52 | .27 | .31 | .46 | .67 |
| High-White 3 | .50 | .51 | .50 | .53 | .44 | .74 |
| Low-Black 3 | .62 | .60 | .58 | .46 | .47 | .77 |

respect to reliability, the next matter of concern is to examine performance on the three as a function of Grades, Populations, and Sex. Since the question of principal interest for each task was whether or not it detected a Population difference at the three Grade levels, the analyses of variance tested the simple main effects of Populations within Grades. For these analyses, the dependent variables for the PPVT, CPM, and PA tests, respectively, were: number of correct responses across item types, lists and trials. The means for these variables are presented in Table 33 and summaries of the three analyses of variance are given in Table 34. Because of the large

---

Insert Tables 33 and 34 about here

---

numbers of hypotheses to be tested in the present study, the probability level for falsely rejecting each null hypothesis was set equal to .01.

As inspection of Table 33 suggests, and the information in Table 34 confirms, for both populations and for all three tests, performance increases as a function of grade level. Note in Table 34 that the proportion of the total between-subjects sums of squares ($\hat{\omega}^2$) associated with Grades seems approximately constant across the three tasks. In contrast, the value of $\hat{\omega}^2$ for Populations appears to vary as a function of both Grades and Tests. In the cases of the PPVT and the CPM, the proportion of the sums of squares attributable to Population differences appears largest for the third-grade samples, whereas in the case of the PA test the proportion seems to be smallest for the third grade and largest for the kindergarten samples. Similarly, the total $\hat{\omega}^2$ for Populations within Grades, summed across grade levels, was larger for the PPVT ($\hat{\omega}^2$=.32) and the CPM ($\hat{\omega}^2$=.28) than for the PA test ($\hat{\omega}^2$=.05). Indeed, among the $\underline{F}$ ratios for Populations, only that for the kindergarten samples was significant on the PA test while $\underline{F}$s were significant for all three grade levels on both the PPVT and the CPM.

With one qualification, the results examined so far are consistent with the hypothesis advanced by Jensen (1969a) concerning Population differences in learning ability. The performance of high-SES Whites substantially exceeded that of low-SES Blacks on both of the tasks (PPVT and CPM) that fall into Jensen's Level II category and the Population difference was considerably less substantial on the task that is presumably of the Level I variety (PA). The qualification, of course, pertains to the fact that the magnitude of Population differences seems to vary considerably with the ages of the $\underline{S}$s sampled. The Level II tasks appeared to yield more variance associated with Populations at the third grade than at the kindergarten level, and even the Level I task revealed a significant Populations difference for the kindergarten samples. Thus, within certain age limits, the present results confirm the first prediction derived from Jensen's hypothesis.

In contrast, the second prediction is clearly disconfirmed by the results of correlational analyses for the three tests administered in the present study. The prediction, it will be recalled, was that the magnitude of the correlation between Level I and Level II tasks should be larger among high-SES White than among low-SES Black $\underline{S}$s. To assess this prediction,

Table 33

Performance on the PPVT, CPM, and PA tests as a Function of Grades, Populations and Sex

| | | PPVT[1] | | | CPM[2] | | | PA Test[3] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Population | Sex | K | 1 | 3 | K | 1 | 3 | K | 1 | 3 |
| High-White | M | 60.6 | 64.5 | 78.9 | 14.3 | 20.5 | 27.1 | 10.2 | 11.7 | 13.4 |
| | F | 58.5 | 63.2 | 76.1 | 15.5 | 19.5 | 26.6 | 8.4 | 10.0 | 13.2 |
| Sub-Total | | 59.6 | 63.8 | 77.5 | 14.9 | 20.0 | 26.8 | 9.3 | 10.8 | 13.2 |
| Low-Black | M | 49.4 | 57.1 | 61.9 | 12.3 | 16.0 | 18.6 | 7.6 | 10.6 | 12.7 |
| | F | 46.0 | 52.2 | 59.2 | 13.2 | 13.8 | 16.1 | 6.7 | 9.4 | 11.4 |
| Sub-Total | | 47.7 | 54.6 | 60.5 | 12.8 | 14.9 | 17.3 | 7.2 | 10.0 | 12.0 |
| Total | | 53.6 | 59.2 | 69.0 | 13.8 | 17.5 | 22.1 | 8.2 | 10.4 | 12.6 |

[1] Mean number of items correct.

[2] Mean number of correct problem solutions.

[3] Mean number of correct responses averaged across trials and lists (maximum possible score = 25).

Table 34

Summaries of Analyses of Variance Performed on Results

Produced by the PPVT, CPM, and PA Tests

| Sources of | | | PPVT | $\hat{\omega}^2$ | | CPM | $\hat{\omega}^2$ | | PA | $\hat{\omega}^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Variance | df | MS | F | | MS | F | | MS | F | |
| Grades | 2 | 5815.1 | 118.1* | .30 | 1641.8 | 110.3* | .31 | 744.7 | 66.9* | .30 |
| Populations/K | 1 | 3384.4 | 68.7* | .09 | 106.3 | 7.1* | .01 | 176.0 | 15.8* | .04 |
| Populations/1 | 1 | 2035.0 | 41.3* | .05 | 620.2 | 41.7* | .06 | 28.4 | 2.5 | --- |
| Populations/3 | 1 | 6936.0 | 140.8* | .18 | 2175.5 | 146.2* | .21 | 58.0 | 5.2 | --- |
| Sex/K | 1 | 176.0 | 3.6 | --- | 25.0 | 1.7 | --- | 70.4 | 6.3 | --- |
| Sex/1 | 1 | 234.4 | 4.8 | --- | 57.0 | 3.8 | --- | 87.6 | 7.9* | .02 |
| Sex/3 | 1 | 181.5 | 3.7 | --- | 55.5 | 3.7 | --- | 25.0 | 2.2 | --- |
| PxS/K | 1 | 9.4 | --- | --- | 0.3 | --- | --- | 7.7 | --- | --- |
| PxS/1 | 1 | 77.0 | 1.6 | --- | 9.3 | --- | --- | 1.8 | --- | --- |
| PxS/3 | 1 | 0.1 | --- | --- | 23.0 | 1.6 | --- | 12.8 | 1.2 | --- |
| Error | 276 | 49.2 | | | 14.9 | | | 11.1 | | |

* $p < .01$

correlation coefficients between the PA test and the PPVT and CPM were calculated separately for each of the six samples. The results are presented in Table 35. An inspection of the values shown in Table 35 indicates that

---

Insert Table 35 about here

---

there is no support in these data for the notion that Level I and Level II abilities are more closely related in high-SES White than in low-SES Black populations at any of the grade levels sampled.

Learning to Learn. For the various purposes attached to the assessment of within *Ss* effects, scores on each of the four lists, five item types and two trials were transformed into 39 new dependent variables and were subjected to multivariate analysis of variance in the manner suggested by Morrison (1967). The interactions of specified within-subjects variables and the between-subjects sources of variation were also examined. With respect to the issue of non-specific transfer or LTL, the variates of interest are the scores obtained on each of the successive four PA lists administered to every *S*. The means for these variables are presented in Figure 31 as a function of Grades and Populations. As an examination of these data suggest,

---

Insert Figure 31 about here

---

the analysis revealed a significant LTL effect, $F$ (3,274) = 26.69, $p$ < .01. A trend analysis confirmed the impression of improvement in performance across lists in that the linear component was significant, step-down $F$ (1,276) = 39.66, $p$ < .01. Although the quadratic component was not significant, step-down $F$ < 1, the uniform drop in performance between the second and third lists administered was detected in the significant cubic component of the trend, step-down $F$ (1,276) = 34.62, $p$ < .01. In this connection it should be recalled that a 48 hour interval elapsed between the administration of the second and third lists. Thus, the drop in performance should probably be attributed to the loss of the benefit of warm-up across the interval. The total amount of improvement in performance from list 1 to list 4, that is, the total amount of nonspecific transfer observed, may be partitioned into two components -- warm-up and LTL. The best estimate of the LTL component is the difference between lists 2 and 4.

With respect to the question of principal interest for the LTL analysis, there was no significant interaction between Populations and practice at any of the three grade levels, all $F$s < 1. Nor were any of the interactions of Sex with practice significant. Accordingly it is warranted to conclude that these data provide no support for the hypothesis that the high-SES White samples profit more than the low-SES Black samples from previous experience with the kind of learning task administered. Indeed, the direction of the differences between second and fourth list performance appears to favor the low-SES Black children (mean difference = 0.97 items), not the high-SES White children (mean difference = 0.10 items).

Table 35

Product-Moment Correlation Coefficients Between Scores on the PA Test and Scores on the PPVT and the CPM as a Function of Grades and Populations

| | Populations | | | |
|---|---|---|---|---|
| | High-SES White | | Low-SES Black | |
| Grades | PPVT | CPM | PPVT | CPM |
| K | .47* | .12 | .66* | .46* |
| 1 | .28 | -.08 | .35 | .02 |
| 3 | .14 | .00 | .38* | .29 |

*$p \leq .01$

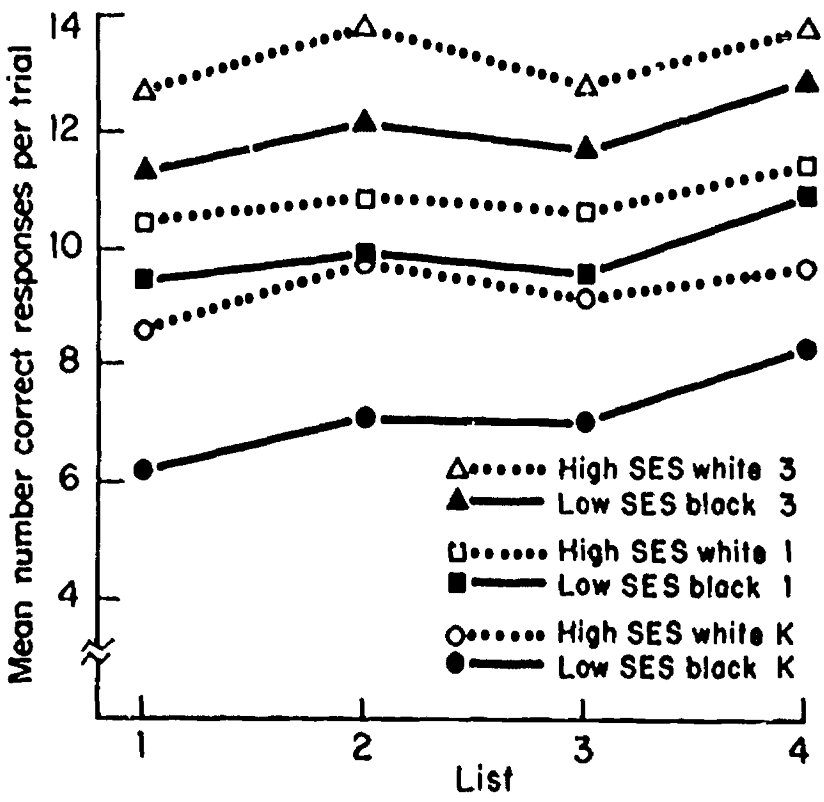



Figure 31. Mean numbers of correct responses on the PA task as a function of grades populations and practice (lists).

Methods of Presentation. Differences among the original Item Type variables are displayed in Figure 32 as a function of Grades and Populations.

---

Insert Figure 32 about here

---

For the multivariate analysis designed to estimate the amount of variance attributable to these differences among methods of presenting the various PAs, four orthogonal linear transformations of the five Item Type variates were made. The effect of Item Types was significant, $F$ (4,273) = 1291.98, $p < .01$. Appropriate post hoc procedures revealed that all pairwise comparisons were significant. Another multivariate test of the transformed variables indicated that the effect of Grades across the five variates was not constant, $F$ (8,546) = 7.44, $p < .01$; the superiority of Picture over Noun items appears to increase with Grades and the superiority of Sentence-Picture and Noun-Action items over Noun, Picture and Noun-Picture items also appears to increase with Grades. It is worth emphasizing the substantial magnitude of the Item Types effect since the estimates of learning efficiency produced by the present study vary so markedly with the method by which the PAs were presented.

With respect to the question whether or not Population differences in learning efficiency depend upon the manner in which PAs are presented, the results indicate that the answer varies with the Grade level sampled. In the Kindergarten samples, the magnitude of the Populations difference varies significantly across Item Types, $F$ (4,273) = 3.52, $p < .01$; descriptively, the effect is that the superiority of the high-SES White sample is greater for Sentence-Picture and Noun-Action items than for Noun, Picture, and Noun-Picture items. Although the Populations effect does not differ significantly across Item Types in the first grade samples, $F$ (4,273) = 2.99, $p = .019$, the direction of the differences appears to indicate that the Populations difference is smaller for the Nouns-Pictures items than for the Nouns and for the Pictures items. A similar pattern of results for Populations across Item Types was detected for the third-grade samples, $F$ (4,273) = 4.24, $p < .01$; Population differences were larger for Nouns and for Pictures items than for Nouns-Pictures items. Indeed, an inspection of Figure 32 reveals that the mean differences on Nouns-Pictures items in both Grade 1 and in Grade 3 favor the low-SES Black samples. In summary, it must be concluded that the detection of Populations differences in PA learning efficiency varies significantly with the method of presentation employed.

Neither the effects associated with the factor of Sex nor those associated with the interaction of Sex and Populations varied significantly across the five Item Types variates. This result holds for all three of the grade levels sampled.

It will be recalled that each list was administered for a total of two trials. An analysis of the transformed variable, Trial 2 score minus Trial 1 score, revealed several interesting effects. This difference itself was significant, $F$ (1,276) = 3454.85, $p < .01$, and the amount of the difference varied with Grades, $F$ (2,276) = 26.09, $p < .01$, such that the gain in correct responses from Trial 1 to Trial 2 increased with grade level. The magnitude of gain also varied significantly as a function of Populations within kindergarten, $F$ (1,276) = 10.68, $p < .01$, and Grade 3, $F$ (1,276) = 8.79,

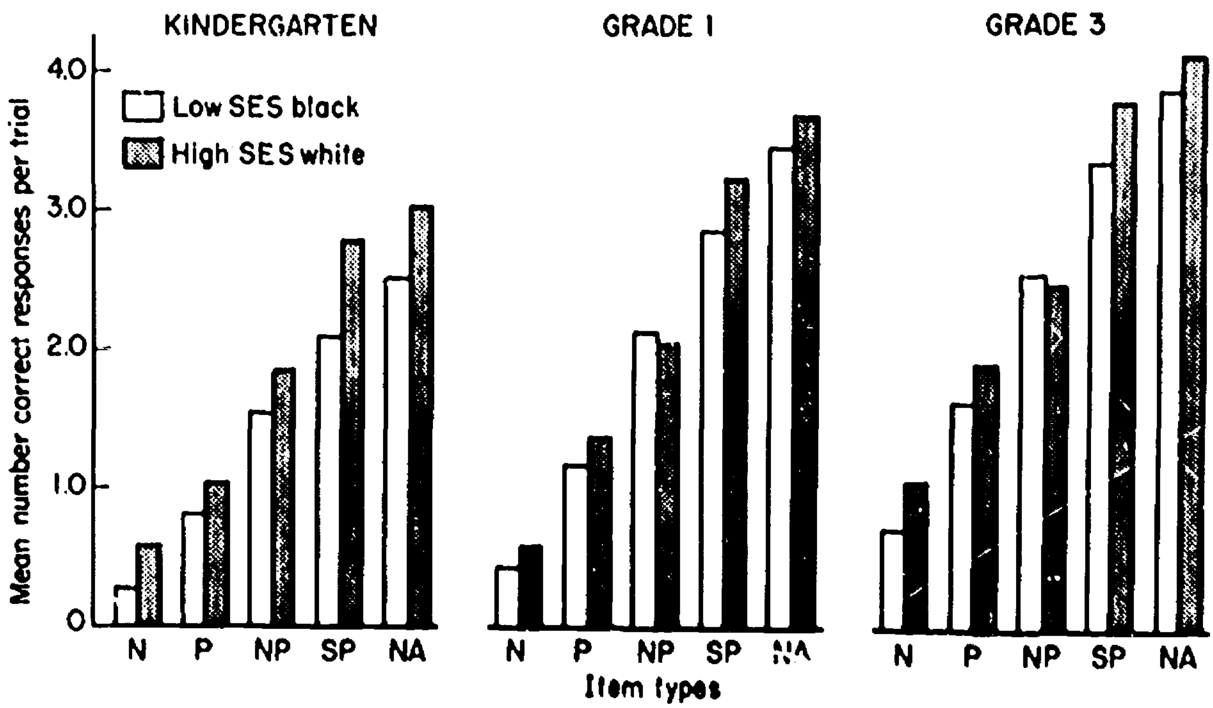



Figure 32. Mean numbers of correct responses on the PA task as a function of grades, populations and item types.

$p < .01$, but not in Grade 1, $F(1,276) = 3.82$, $p = .052$. In the case of each of the significant effects, the high-SES White samples appear to gain more from trial to trial than the low-SES Black samples. Trials did not differ significantly as a function of the factor of Sex or as a function of the interaction of Populations and Sex at any of the three Grade levels.

Retention. The efficiency of recall for the first two PA lists administered after a two-day retention interval was indexed in two ways: number of correct responses given on the recall trials; and amount lost between the second test trial of original learning and the recall trial, that is, number of correct responses given on Trial 2, Day 1, minus number of correct responses on the recall trial, Day 2. The results for each measure are presented in Table 36 as a function of Grades and Populations. Analyses of variance were performed on both measures of retention but only that for

- - - - - - - - - - - - - - -

Insert Table 36 about here

- - - - - - - - - - - - - - -

the variable of amount lost will be reported, since the variable of number of items retained reflects primarily the efficiency of original learning rather than the efficiency of recall. The main effect of Gradea was significant, $F(2,276) = 31.06$, $p < .01$, such that fewer items were lost by kindergarten than by first- and third-grade children; the means for the first- and third-grade samples did not differ significantly. The effect of Populations was significant for the kindergarten samples, $F(1,276) = 11.03$, $p < .01$, but not for Grade 1, $F(1,276) = 2.03$, $p > .05$, or for Grade 3, $F < 1$. As can be seen in Table 36, there is evidence of more forgetting on the part of the high-SES White kindergarten children than on the part of the low-SES Black children. Clearly, there is no evidence in theae data to support the supposition that low-SES Black children are deficient in their capacity for retaining what they learn.

The amount of variance in forgetting associated with Sex was not significant for any of the three Grade levels. The interaction of Populations with Sex, however, was significant for the kindergarten samples, $F(1,276) = 8.63$, $p < .01$, and not for the other two Grades, both $F$s $< 1$. Descriptively, in the kindergarten samples the form of the interaction is such that for high-SES Whites, more items were lost by males than by females, whereas for low-SES Blacks, more items were lost by females than by males.

The decreases in number of correct responses from Trial 2 of original learning to the recall trial for each of the Item Types were transformed into four new variables. The multivariate test for equality of these decreases was significant, $F(4,273) = 115.29$, $p < .01$. The mean decreases for each original variate were: Nouns, .67; Pictures, 1.13; Nouns-Pictures, 1.78; Sentences-Pictures, 1.58; Nouns-Action, 1.77. Appropriate post hoc procedures revealed no significant differences among the Nouns-Pictures, Sentences-Pictures, and Nouns-Action Item Types (although each of these differed significantly from both Nouns and Pictures). This outcome is important in connection with the problem presented by the generally positive correlation between the variable of amount learned and that of amount lost. This pattern is broken in the present analysis where no more pairs were lost in the Sentences-Pictures than in the Nouns-Pictures Item Types, even though

Table 36

Mean Number of Items Recalled on Day 2 and Mean Number of Items Lost by Day 2 as a Function of Grades and Populations

| | Mean Number Recalled | | Mean Number Lost | |
|---|---|---|---|---|
| Grades | High-SES White | Low-SES Black | High-SES White | Low-SES Black |
| Kindergarten | 5.45 | 4.05 | 6.25 | 4.62 |
| First Grade | 6.00 | 5.45 | 7.61 | 6.92 |
| Third Grade | 8.10 | 6.60 | 8.34 | 7.86 |

the number of correct responses during Trial 2 of original learning was greater for the former than for the latter types.

No one of the multivariate tests for Item Types by Populations within Grades was significant: Kindergarten, $\underline{F} < 1$; Grade 1, $\underline{F} < 1$; Grade 3, $\underline{F}(4,273) = 3.06$, $\underline{p} = .017$. Nevertheless, it may be useful to examine in more detail the Item Types results relevant to Population differences because of the generally positive correlation between amount originally learned and amount lost. The means for these comparisons are depicted in Figure 33.

Insert Figure 33 about here

The evidence relevant to the issue of retention differences would be clearer if recall comparisons could be located where the two Populations were equivalent at the end of original learning. An inspection of Figure 32 reveals two such instances: performance on Nouns-Pictures items in the first- and third-grade samples. (Although Figure 32 displays mean performance across Trials, it accurately reflects Trial 2 performance as well.) If low-SES Black children are in fact deficient in their capacity to recall material previously learned, despite the overall results of the present study to the contrary, this deficiency should exhibit itself in higher mean loss scores for the first- and third-grade samples on the Nouns-Pictures Item Type. The data displayed in Figure 33,however, offer no support whatever for this hypothesis. Accordingly, for the task used here, it must be concluded that low-SES Black children do not show a deficiency in relation to high-SES White children in their capacity to retain what they have learned.

Finally, none of the Item Types tests for the effects of Sex or for the interactions Populations x Sex was significant.

## Discussion

The purpose of the present investigation was to establish some facts necessary for evaluating an explanation of the observed discrepancy in school achievement between high-SES White and low-SES Black children. At issue is the question whether or not this discrepancy can be accounted for in terms of a corresponding discrepancy in learning proficiency. If learning proficiency is indexed by instruments of the intelligence test variety, the relative performance of the two populations in the present study on the PPVT and the CPM lend support to the explanation. In contrast, if learning proficiency is indexed by a task that directly engages $\underline{S}$s in learning, in this case the PA test, the relative performance of the two populations contradicts the explanation. Thus, the issue remains unresolved; one method for indexing learning proficiency reveals a Populations discrepancy consistent with that observed for school achievement while another method finds no such discrepancy.

In examining some attempts to reconcile these disparate outcomes, it will be useful to recall the present results in detail. The data clearly indicate that the PA test yields scores approximately equivalent in reliability to those yielded by the PPVT and CPM. The results also confirm

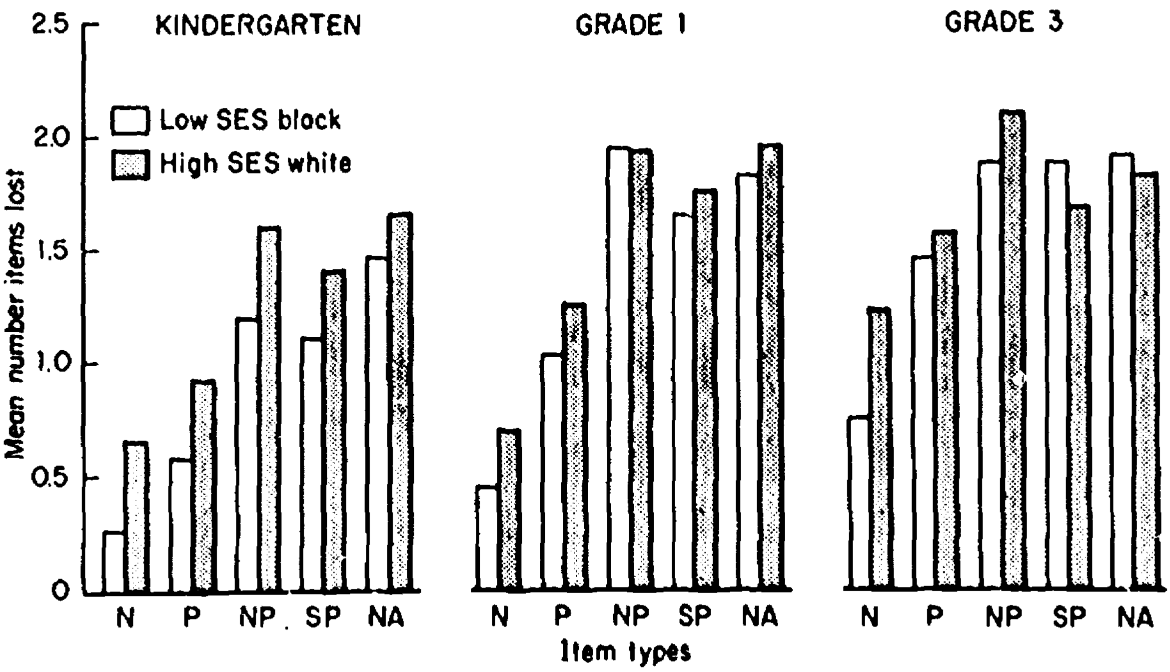



Figure 33. Retention of PAs (mean number of items lost) as a function of grades, populations and item types.

the inference from previous research that the magnitude of populations differences depends on both task variables and chronological age. On the PPVT and the CPM, the performance of high-SES White children exceeded that of low-SES Black children at all grade levels sampled, kindergarten, first and third, but the amount of variance associated with the populations difference appeared larger for third-grade than for kindergarten children. On the PA test, however, the Populations difference was significant only for the kindergarten samples. Again with respect to the PA test, when interactions with Item Types are ignored, low-SES Black children are not deficient relative to high-SES White children in either the efficiency of original learning or in the retention of what has been learned or in amount of nonspecific transfer, that is, in the degree to which they benefit from previous learning. Thus, the problem is to provide an account of Populations differences in school achievement that is consistent with the results produced both by the PA test and by the PPVT and the CPM.

The hypothesis advanced by Jensen (1969a) is an attempt to give precisely this kind of account. In brief, the relevant rationale that might be derived from the Level I - Level II model is as follows. There are Populations differences in the abilities necessary for successful performance on Level II tasks, here exemplified by the PPVT and CPM; the abilities required by Level I tasks (e.g., the PA test), that is, associative abilities, are distributed equally across populations. Because of the character of instruction, Level II abilities are required for successful performance on school learning tasks. Therefore, high-SES Whites perform better on school achievement tests than low-SES Blacks. In this fashion, the Level I - Level II model can provide a reconciliation for the disparate results presented here while simultaneously accounting for Populations differences in school achievement.

Considerable caution in accepting this interpretation, however, is warranted by two features of the present study. The first concerns the assumption that the PA test mainly elicits processes of the Level I or associative variety. As Rohwer (in press) has argued, this assumption is questionable in view of the evidence that PA learning involves considerable conceptual activity (Bower, 1968, 1969; Bugelski, 1962; Martin, 1967; Martin, Boersma & Cox, 1965; Montague & Wearing, 1967; Paivio, 1967, Paivio, Yuille & Smythe, 1966; Rohwer, 1967, 1970; Rohwer & Levin, 1968; Rohwer & Lynch, 1966, 1967; Rohwer, Lynch, Levin & Suzuki, 1967, 1968; Rohwer, Lynch, Suzuki & Levin, 1967; Rohwer, Shuell & Levin, 1967; Runquist & Farley, 1964). To be sure, PA tasks require verbatim reproduction of responses, but this characterization of the performance demanded does not necessarily imply that PAs are learned by rote processes. Another reason for caution in accepting the Level I - Level II account is raised by the results of the correlational analyses presented here. It will be recalled that from the model, Jensen (1969a) derives the prediction that performance on Level I tasks will be more highly correlated with performance on Level II tasks among high- than among low-SES children. If the designations of the PPVT and the CPM as Level II tasks and of the PA test as a Level I test are accepted, then the correlational results presented here run directly counter to the prediction. Thus, on both counts the Level I - Level II model fails to provide a satisfactory explanation of Populations differences in school achievement.

Among alternative ways of accounting for this phenomenon, at least two are quite obvious possibilities. One is that the reasons for the observed deficiencies of low-SES Black children in performance on school achievement and IQ tests are not to be found in the domain of cognitive or intellective variables. This is conceivable, but on its face, it seems unlikely that some substantial portion of the variance in performance on intellectual tasks cannot be accounted for in terms of intellective variables. The second obvious possibility is that PA tasks do not elicit the kinds of learning processes necessary for successful performance on school learning tasks or on IQ tests. Clearly, this answer is subject to empirical evaluation; one method of assessing it is to compare the validity of the three tests administered in the present investigation for predicting performance on school achievement tests. Such a study is presently in progress. Meanwhile, caution should be exercised in concluding that PA learning is irrelevant to school learning in view of the substantial relationships reported by Stevenson, Hale, Klein and Miller (1968) between PA learning and school achievement.

If neither of these obvious possibilities is satisfactory, it is in order to consider still another alternative. One such has been proposed by Rohwer (1969). It begins by noting that any instrument which demands that the testee recall previous learnings will inevitably reveal Populations differences unless the degree of original learning has been equivalent among the populations assessed. If it is granted that high-SES White children achieve higher degrees of mastery than low-SES Black children by the end of any given instructional unit, then any tests that probe for the recall of material learned from that unit will show a Populations difference. This category of instruments would include the PPVT and virtually any standardized test of school achievement. This analysis, however, does not account for the fact that Populations differences have been observed on some tasks that principally require new learning as well as on tasks that require the recall of previous learnings. One example is provided by the results of the CPM presented here; other examples include performance on tasks as straightforward as digit span and the free recall of lists of familiar objects that are subsumable in formal categories. Rohwer (1969) has suggested that Populations differences on tasks such as these are attributable to a common property: efficient performance on each task requires the mastery of sets of formal conventions (for example, numbers and categories) created by cultural consensus that may be more readily available to or more valued by one population than by another. One implication of this position is that tasks where proficient performance depends more on skill at the application of imaginative, idiosyncratic conceptual processes than upon conventional formal processes will reveal equivalence of learning efficiency.

This hypothesis directs attention to the conditions of original learning, for it is these conditions, to a large extent, that determine the degree of learning that will be achieved and the manner of its achievement. In this connection, the results of the present study are provocative in three respects. First, they suggest the capacity of low-SES Black children to recall previous learnings is as great as that of high-SES White children. Second, if it is assumed that successful performance on PA tasks involves the operation of imaginative conceptual processes, the results are consistent with the expectation that Populations equivalence should be observed. Third, they make it clear that a truly remarkable amount of variance in the success of original learning, for both Populations, is associated with Item Types, that is, with the manner in which the learning materials are presented.

It should be emphasized that in the hypothetical account adapted from Rohwer (1969) and offered here, substantial weight is given to those variations in the present results that were associated with Item Types differences. Accordingly, it is important to ascertain the generality of the effects of these differences and to discount the possibility that they are an artifact of the specific procedures followed. In particular, it should be determined whether the Item Types effects are confined to the mixed list design or if they also hold true when independent groups designs are used. There is some indication that the Item Types effect is general to the independent groups case (Rohwer, Lynch, Levin & Suzuki, 1968), but a direct comparison of the two methods has not yet been made.

Finally, the speculative character of this account of Populations differences in school achievement must be made explicit. The distinction between tasks that elicit formal as against imaginative conceptual processes lacks clarity in the sense that operations for distinguishing among tasks have not been specified. Thus far, the account is an *ad hoc* one, relying largely on the results of the present study and of other similar studies using PA tasks for its empirical support. Accordingly, it should be evaluated with respect to other kinds of tasks rather than only in terms of the tasks that spawned it.

Elaboration Training and Paired-Associate Learning

Efficiency in Children[1]

William D. Rohwer, Jr. and Mary Sue Ammon

The present study was designed to evaluate an attempt to increase, through training, paired-associate (PA) learning efficiency in children. Such an attempt is of interest from two perspectives, one psychological, the other educational. In the psychological domain, several recent experiments indicate that substantial increments in PA learning efficiency can be produced by the elaboration of the pairs to be learned. The specific meanings of the term elaboration can be illustrated in connection with a particular task, that of learning a list of noun pairs. In this kind of task, elaboration may be directed at either the individual items that comprise each pair or at the pair unit itself. Consider first the case of elaborating individual items. If the materials are presented aurally, elaboration consists of representing the referents of each noun visually as in a picture or, more loosely speaking, in an image; if the materials are presented pictorially, elaboration consists of representing each item verbally as by the appropriate noun label for the given object. Second, consider the case where both members of each pair are included in a single elaborative unit. As in single-item elaboration, pair elaboration may be either verbal or pictorial: the two nouns can be used to form the subject and object of a sentence describing an event; or, an event involving the two objects named by the nouns can be depicted pictorially.

Several strands of evidence support the presumption that each of these four forms of elaboration can increase the efficiency of learning noun pairs. With respect to single-item elaboration, it has been demonstrated repeatedly (cf. Paivio, 1969) that the learning of high-imagery value noun pairs is more efficient than the learning of low-imagery value noun pairs. Similarly, Rohwer, Ammon, Suzuki and Levin (in press) have reported that PA performance is better when noun labels are presented concurrently with pictorially presented object pairs than when the pictures are presented alone.

The positive effect of pair elaboration on learning efficiency has been documented by means of three different methodologies: post-learning interviews, pre-learning instructions to elaborate, and manipulation of the conditions of presentation. The post-learning interview technique has revealed: (a) that subjects report elaborating PAs by constructing sentences containing the word pairs or by forming images involving the referents of the pairs (e.g., Bugelski, 1962; Runquist & Farley, 1964); (b) that the kinds of elaboration reported

[1]We wish to acknowledge the indispensible assistance of the teachers and children in those schools where the data were collected as well as the contributions of Dr. Nancy Suzuki in administering the various tests and of Dr. Joel R. Levin in analyzing the results.

can be classified reliably with respect to complexity (Martin, 1967); and, (c) that there is a positive correlation between complexity of the elaboration reported and the efficiency of PA learning (Martin, 1967; Montague & Wearing, 1967). The efficiency of PA learning has also been shown to vary as a function of whether or not the pairs are presented in an elaborated form. In this connection, both verbal and visual forms of elaboration have proved to be effective; performance is increased by presenting noun pairs in the context of sentences or relational phrases (Davidson & Adams, 1969; Rohwer, 1967) or by depicting the referents of the nouns in a pictorial interaction (Milgram, 1967b; Reese, 1965; Rohwer, Lynch, Levin & Suzuki, 1967). Finally, instructions to elaborate noun pairs either by forming sentences or images have been observed to produce substantial increments in performance (Jensen & Rohwer, 1963, 1965; Bower, 1969; Milgram, 1967a).

Since brief elaboration instructions have demonstrable effects on PA performance when given immediately before the onset of the learning task, more extensive training in elaboration skills should produce more enduring effects. It was expected that such training would result in detectable positive transfer in performance on PA lists administered outside of the context in which the training was provided. Thus the primary question of psychological interest was whether or not elaboration training can be shown to make durable differences in the efficiency with which children learn PA tasks.

From an educational perspective, two features of the present study were prominent. The first was whether elaboration training improves PA performance more than simple practice on PA tasks for equal amounts of time. The second concerned the issue whether or not elaboration training would suffice to reduce observed discrepancies in learning efficiency between children classified as low-SES Black and children classified as high-SES White. Although the differences between these two populations are much smaller on PA tasks than on intelligence and achievement tests, they are frequently detected among young children, that is, eight years of age and under. In particular, they are detected when the method of presenting PAs does not provide <u>item</u> elaboration and when it does provide <u>pair</u> elaboration (Rohwer, Ammon, Suzuki & Levin, in press). Accordingly, the PA tasks used to evaluate the effects of elaboration training were constructed to permit an evaluation of this interaction between populations and methods of presentation.

## Method

<u>Subjects</u>. Sixty children were randomly selected from the second-grade classes in each of two public elementary schools. One of the schools serves a low-SES Black residential area and the other serves a high-SES White area. The SES designations were based on average census tract information collected in the 1960 survey.

<u>Design</u>. The factors in the 2 x 3 x 20 design were Populations (high-SES White vs. low-SES Black), Treatments (Training, Practice, Control), and Levels. Assignment to levels was determined by pretest

performance; within each population, the sample was subdivided into twenty groups of three Ss each in terms of pretest similarity and the members of each level were randomly assigned to treatment conditions.

The three treatments were distinguished by the activities prescribed for the pretest-posttest interval. In the Training condition, instruction was designed to facilitate the acquisition of the following skills in connection with PA learning tasks: (a) self repetition of presented pairs; (b) verbal labelling of objects presented pictorially; (c) visualization of noun referents presented aurally; (d) the visualization of action sequences involving object pairs; (e) the generation of sentence descriptions of such action sequences. Instruction relating to these skills was conducted in connection with several PA lists comprised of a variety of types of items -- objects, pictures, aurally presented nouns. The following are examples of training activities. The trainer presented an array of three-dimensional toy objects and asked the child to select any two and to perform an action episode involving them. The child was then asked to describe the performance or "story" verbally. In another activity, pictures of object pairs were presented on a television monitor; the child was directed to visualize the picture story involving the two objects and to give a verbal description of the story or to guide the trainer in making a drawing of the episode.[1]

Instruction in the Practice condition was directed only at the skill of self repetition of presented pairs. Thus the major activity in the training sessions for this condition was the learning of successive PA lists.

Certain other features of these two treatment conditions were common to both. In both content and presentation mode, the PA lists used for instructional purposes in the Training condition were identical with those used in the Practice condition. A variety of presumed incentives was made available to children in both conditions: a chart was made for each child which displayed a record of performance improvement as a function of sessions; colorful stickers were awarded for successful performance; and, the child was allowed to play with hand puppets after performing on a PA list. In sum, children in both conditions received equivalent exposure to the same kind and number of PAs and received the same kinds of feedback in connection with their performance. Thus, the structure of the Practice condition was designed to control for the effects of all experiences provided in the Training condition except those involving direct instruction in elaboration techniques.

In contrast, children assigned to the Control condition received only the pretest and the posttest.

Materials. Both the pretest and the posttest consisted of two different kinds of 25-item lists of noun pairs. In each case the two

---

[1]A detailed script used in conducting the training sessions may be obtained from the first author on request.

lists differed with respect to the manner in which the items were presented: in the one (aural) all pairs and test stimuli were presented aurally by means of an audio tape recorder; in the other (mixed) equal numbers of items were presented in each of five different ways. Thus the mixed PA list yielded scores on five Item Types: noun pairs presented aurally (Nouns); photographs of object pairs depicting noun referents (Pictures); photographs of similar objects with their noun labels presented aurally (Nouns-Pictures); photographs of object pairs with their noun labels presented aurally in the context of a sentence (Sentences-Pictures); photographs of an event involving the members of object pairs with their noun labels presented aurally (Nouns-Action). For the mixed list, the test stimuli consisted of one of the items from each pair presented in the same manner as on the study trials. For example, test stimuli for pairs of the Pictures type were photographs of single objects from each of the relevant pairs. All materials for the mixed list tests were recorded on videotape and presented by means of a television monitor.

For both the mixed and the aural lists, the pre- and posttests were administered for two study and two test trials each. In the case of both kinds of lists, the rate of presentation on study and on test trials was four and one-half seconds and the intertrial intervals were four and one-half seconds. The order of items was different on every trial.

Two considerations led to the choice of the particular PA lists used as the pre- and posttests. The first pertains to the mixed lists which were selected to permit an assessment of the differential effects of training as a function of presentation methods. Furthermore, these lists had been used previously in such a way as to yield both reliability and difficulty estimates for children from populations similar to those sampled in the present study. The consideration leading to the use of the aural lists was that of presumed sensitivity to training effects. A list of 25 aurally presented PAs is more difficult than a mixed list of comparable length. Accordingly, it was reasoned that the more difficult aural list might detect training effects that would not appear in the easier mixed list.

Procedure. The study was conducted by two female Es, one of whom administered all pre- and posttests while the other administered all of the treatment sessions. The E responsible for the pre- and posttests did not participate in the assignment of Ss to treatment conditions nor was this information available to her at any time during the study.

The major events of the study consisted of the administration of the two pretest lists, the assignment of Ss to levels, the assignment of Ss within levels to treatment conditions, the administration of the various treatments, and the administration of the posttest lists.

The two posttest lists were given to Ss individually. The order of administration was constant across all Ss: the mixed list followed by the aural list. Prior to the onset of each list, instructions were given informing the S of the nature of the materials to be presented and of the task he was expected to perform. These instructions were illustrated by five sample items presented in accord with the same study-test pattern as that which characterized the test lists. The sample items were repeated until the S responded correctly to three of the five test stimuli in the list. Following the attainment of this criterion, the S was told that the learning task proper would begin and the 25-item list was administered.

For the purpose of assignment to levels and to treatment conditions, two scores were calculated for each S: total number of correct responses on the two test trials of the mixed list, and total number of correct responses on the two test trials of the aural list. Within each of the two populations sampled, Ss were then grouped into sets of three such that the members of each set were characterized by similar scores on both the mixed and the aural lists. Thus, Ss having high scores on both lists were grouped together, those having low scores on both were grouped together and those having high scores on the mixed and low scores on the aural list were grouped together and vice versa. This procedure yielded 20 levels of three Ss each. Within each level, Ss were randomly assigned to the three treatment conditions.

The method of assigning Ss to levels in terms of their scores on the two pretest lists considered separately permitted the application of a randomized analysis of variance design to the results yielded by each of the two posttest lists. It should be noted that the same rank order of levels did not obtain across the analyses for the two posttests. For example, the three Ss assigned to the highest ranking level for the mixed list analysis may have occupied a somewhat lower ranking level for the analysis applied to the results of the aural list. Nevertheless, the various sets of three Ss remained intact across both analyses even though the rank of the sets among the levels varied from one analysis to another.

The Training and Practice treatments were administered in five sessions of 20 to 30 minutes duration each on five successive school days. Two Ss were given the assigned treatment sessions together rather than singly in order to conserve time. The treatment sessions for the high-SES White sample were conducted during the first and third weeks of this phase of the study and those for the low-SES Black sample were conducted during the second and fourth weeks. For each S a record was made of the number and the identity of the treatment sessions missed in order to permit a post assessment of the relationship between attendance and posttest performance. In all cases, sessions missed were attributable to absence from school.

The schedule developed for the posttest phase specified that equal numbers of Ss from each of the three treatment conditions would be administered the final mixed and aural lists during the week immediately following the treatment sessions and the administration of the posttest. The procedure followed in giving the posttest lists was the same as that described for the pretest lists.

## Results

Pretest performance was used to assign Ss to levels within Populations. Accordingly, the effects of the treatment conditions were assessed entirely in terms of performance on the posttests, aural and mixed list.

Aural posttest. The results yielded by the aural posttest are shown in Table 37 as a function of Populations and Treatments. The data were

- - - - - - - - - - - - - - -
Insert Table 37 about here
- - - - - - - - - - - - - - -

evaluated by analysis of variance in which the sources assessed were: Populations, Levels nested within Populations, and the simple main effects of Treatments within Populations; the variance associated with Treatments was partitioned into components to yield two planned comparisons: Practice with Control, and the average of Practice plus Control with Training. In the case of the latter comparison, a directional prediction favoring the Training condition was made in advance thus calling for a one-tailed test; the remaining tests were two-tailed and all were made with $\alpha = .05$.

As the means in Table 37 suggest, the Populations difference was significant, $F\ (1,74) = 27.58$, $p < .01$;[1] the high-SES White Ss scored higher than the low-SES Black Ss. The main effect of Levels was significant for the high-SES Whites, $F\ (19,74) = 3.13$, $p < .01$, but not for the low-SES Blacks, $F < 1$. The difference in the magnitude of these two F ratios may indicate that performance on the aural list was more stable between the pre- and posttests for the high-SES Whites than for the low-SES Blacks.

With regard to the question of principal interest, the analysis of variance revealed quite different effects of Treatments for the two populations. For the high-SES Whites, the difference between the Practice and Control conditions was not significant ($F < 1$) whereas performance in the Training condition was superior to the average of those two conditions, $F\ (1,74) = 6.31$, $p < .01$. Practice on the task of learning PA lists had no detectable effect on posttest performance on the aural list but the instruction provided in the Training condition was successful in improving performance on this same criterion test. The results for the low-SES Black samples were not at all comparable: the F ratios for both planned comparisons were less than 1. Neither practice in PA learning nor specific training in elaboration skills resulted in performance increments detectable on the aural posttest.

Mixed-list posttest. The results of the study as indexed by the mixed-list posttest are presented in Table 37 as a function of Populations and Treatments. For the between subjects factors, the analysis design applied to the mixed-list posttest was the same as that described

---

[1] In all analyses reported, observations were estimated for 2 Ss lost by virtue of absence from 4 or 5 training sessions and during the entire posttest period. Therefore all relevant dfs have been adjusted accordingly.

Table 37

Mean Numbers of Correct Responses on the Aural List and Mixed List Posttests as a Function of Populations and Treatments

Aural List

| Populations | Training | Practice | Control | Total |
|---|---|---|---|---|
| High-SES White | 5.95 | 4.68 | 3.92 | 4.85 |
| Low-SES Black | 2.68 | 2.40 | 2.58 | 2.55 |

Maximum Score = 25

MSE (74) = 11.51

Mixed List

| Population | Training | Practice | Control | Total |
|---|---|---|---|---|
| High-SES White | 14.02 | 14.15 | 12.18 | 13.45 |
| Low-SES Black | 12.98 | 11.58 | 11.62 | 12.05 |

Maximum Score = 25

MSE (74) = 13.95

for the aural posttest. The main effect of Populations was significant once again, $F$ (1,74) = 8.33, $p$ < .01, with more correct responses observed for high-SES Whites than for low-SES Blacks. The variance associated with Levels was significant in both Populations, $F$s (19,74) = 2.96 and 2.74, $p$ < .01, for the high-SES Whites and the low-SES Blacks respectively. Thus, in contrast to the outcome on the aural posttest, performance seems to have been of equivalent degrees of stability for both populations on the mixed-list posttest.

An inspection of Table 37 suggests and the analysis confirms that performance in the Practice condition for the high-SES Whites was significantly superior to that in the Control condition, $F$ (1,74) = 5.59, $p$ < .05. Indeed, descriptively speaking, the Training and Practice conditions differ only negligibly and both may be presumed superior to the Control. Accordingly, it may be inferred that for this sample, Training had no specific effects beyond those produced by simple practice on PA tasks. It is consistent with this interpretation that the variance associated with the comparison between the Training condition and the average of the Practice and Control conditions was not significant, $F$ (1,74) = 1.42, $p$ > .05. These results are in marked contrast to those observed for the low-SES Black sample. For this Population, the Practice condition afforded no advantage over the Control, $F$ < 1, but the Training condition was superior to the average of Practice and Control, $F$ (1,74) = 3.61, $p$ < .05.

In summary, the analyses of between-subjects sources of variance reveal that the effects of training depend on both the character of the posttest list and the population sampled; the Training condition produced specific facilitative effects for the high-SES Whites only on the aural posttest, not on the mixed-list posttest, whereas for the low-SES Blacks, the Training condition facilitated learning on the mixed-list posttest but not on the aural list.

It will be recalled that the mixed-list posttest was comprised of five different types of items: Nouns, Pictures, Nouns-Pictures, Sentences-Pictures, and Nouns-Action. For the multivariate analysis designed to estimate the amount of variance attributable to these differences among methods of presenting the various PAs, four othogonal line transformations of the five original Item Type variates were made. The effect of Item Types was significant, $F$ (4,71) = 391.63, $p$ < .01. Another multivariate test of the transformed variables revealed that the effect of Populations was not the same across all of the variates, $F$ (4,71) = 5.36, $p$ < .01. Differences among the original Item Types variables that are relevant to these tests are displayed in Table 38 as a function of Populations.

- - - - - - - - - - - - - - -

Insert Table 38 about here

- - - - - - - - - - - - - - -

An inspection of Table 38 suggests that the Populations difference may have been confined to the Item Types, Nouns-Pictures, Sentences-Pictures, and Nouns-Action; descriptively, the differences on the Noun Item Type and the Pictures Item Type appear negligible. Since this

Table 38

Mean Numbers of Correct Responses on the Mixed-List Posttest and on the Mixed-List Pretest as a Function of Populations and Item Types

Posttest (n=118)

| Populations | Nouns | Pictures | Nouns-Pictures | Sentences-Pictures | Nouns-Action |
|---|---|---|---|---|---|
| | | | Item Types | | |
| High-SES White | 0.82 | 2.66 | 2.52 | 2.84 | 4.62 |
| Low-SES Black | 0.72 | 2.63 | 1.89 | 2.57 | 4.25 |

Pretest (n=120)

| Populations | Nouns | Pictures | Nouns-Pictures | Sentences-Pictures | Nouns-Action |
|---|---|---|---|---|---|
| | | | Item Types | | |
| High-SES White | 0.69 | 1.08 | 2.48 | 3.02 | 4.02 |
| Low-SES Black | 0.38 | 0.45 | 2.54 | 2.45 | 3.42 |

pattern of results across Item Types varied substantially from that reported previously (Rohwer, Ammon, Suzuki & Levin, in press), an examination was made of performance on the pretest. The relevant results are also presented in Table 38 as a function of Populations and Item Types.

It must be emphasized that the pretest analysis included the scores of subjects later lost from the training study for absenteeism so that there are no meaningful direct comparisons between the results for the two tests. Nevertheless, the difference in the pattern of correct responses across Item Types is interesting in three respects. First, note that the Item Type showing the greatest Populations difference on the posttest, namely, Nouns-Pictures, revealed virtually no Populations difference on the pretest. Secondly, on the pretest, the two populations appeared to differ on all of the remaining four Item Types, including those showing virtually no difference on the posttest, that is, Nouns and especially Pictures. Finally, note that the Nouns-Picture Item Type on the pretest was associated with more correct responses than the Pictures Type, pronouncedly so for the low-SES Blacks, whereas on the posttest, the Pictures Item Type was superior to Nouns-Pictures for both Populations. To be sure this inconsistency between pre- and posttest performance may be attributable to list differences. The fact, however, that both these lists, the pretest list and the posttest list, have produced quite consistent patterns of results in previous use (Rohwer, Ammon, Suzuki & Levin, in press) does not encourage adoption of this interpretation.

The multivariate tests of most pertinence to the present investigation, of course, are those appropriate for determining whether the effects of the Treatment conditions were constant across Item Types in the two Populations. This description, constant effects across Item Types, is supported by the results of the four relevant multivariate tests; for the high-SES Whites and the low-SES Blacks respectively, neither the variance associated with the comparison of Practice with Control, $\underline{F}s\ (4,71) = 1.50$ and $1.61$, $\underline{p} > .05$, nor that associated with Training vs. the average of Practice and Control, $\underline{F}s\ (4,71) = 0.35$ and $1.95$, $\underline{p} > .05$, was significant. Thus, it is warranted to restate the conclusion that specific elaboration training is effective for high-SES White children on the aural posttest and for the low-SES Black children on the mixed-list posttest.

## Discussion

Four aspects of the present results call for additional comment. The most obvious of these is that elaboration training can, in fact, increase the learning efficiency of high-SES White and low-SES Black second-grade children. A second interesting feature of the results is the specificity of the training effect with respect to Populations and type of posttest. The content of instruction in the Training condition was omnibus in character, involving both visual and verbal skills at both the item and the pair level. Nevertheless, the effects of training were not general. For the high-SES White children, the training effects were discernible only on the more difficult posttest task, the aural list, where neither item nor pair elaboration was provided in the materials.

This result is comprehensible in terms of the notion that the aural list provided a maximum opportunity for the operation of subject-initialed elaborative activity. The problem with this account is that the aural list detected no training effect among the low-SES Black children; for this sample, the Training condition was superior to the Practice and Control conditions only on the mixed-list posttest. The one clear inference from this pattern of results is that the effect of training, although positive in both cases, was different for the two Populations. It is not possible to discern the specific character of this difference in the present results. Accordingly, the results lend force to the recommendation that in training studies posttest tasks should be selected which can resolve general treatment effects into their components and which elicit a variety of performances sufficient to accommodate the possibility that training may have different effects for different populations.

A third noteworthy aspect of the present results is the Populations difference itself. One of the objectives of the Training treatment was to remediate deficiencies in the elaborative learning skills of the low-SES Black children relative to the high-SES Whites. The results indicate that this attempt was, in a practical sense, a failure; that is, on both posttest tasks, the learning efficiency of the low-SES Blacks approximated that of the high-SES Whites more closely in the Control condition than in the Training condition. This outcome is not surprising in the case of the aural list since it detected no effect of training for the low-SES Black sample but the same pattern is evident on the mixed list which did show a positive effect of training. Thus, implementation of the Training treatment used in the present study should be made with full recognition that its likely effect will be beneficial for both Populations but more so for high-SES Whites than for low-SES Blacks.

Finally, still another cautionary note must be sounded about the potential of instruction in elaborative skills for attaining the objective of improving learning efficiency in second-grade children. The data presented in Tables 37 and 38 suggest that the method of presenting learning materials can have a more pronounced effect on learning efficiency in young children than the kind of training experience which they are given; the differences among Item Types appear larger than the differences among Treatment conditions. It may be speculated that for children in the age range sampled here, the acquisition of designated content can be better assured by careful design of the method of presentation than by attempts to improve the learning skills of the child. It could be that training in elaborative skills would be more productive for older than for younger children.

## References

Bloom, B. S. Stability and change in human characteristics. New York: Wiley, 1964.

Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and of sequential constancies in free recall. Psychological Reports, 1966, 19, 935-942.

Bower, G. H. Mental imagery and memory. Paper presented at several psychology departments. 1967.

Bower, G. H. Organizational factors in memory. Research Seminar, Institute of Human Learning, University of California, Berkeley, 1968.

Bower, G. H. Mental imagery and associative learning. Paper presented for the Fifth Annual Symposium on Cognition, Carnegie-Mellon University, April, 1969.

Bruner, J. S. On cognitive growth: I, II. In J. S. Bruner, R. R. Olver, & P. M. Greenfield, et al., Studies in cognitive growth. New York: Wiley, 1966.

Bugelski, B. R. Presentation time, total time, and mediation in paired-associate learning. Journal of Experimental Psychology, 1962, 63, 409-412.

Coleman, J. S., et al. Equality of educational opportunity. U. S. Department of Health, Education, and Welfare, 1966.

Davidson, R. E., & Adams, J. F. Verbal vs. pictorial facilitation of paired-associate learning. Paper presented at the annual meeting of the American Educational Research Association, 1969, Los Angeles.

Dilley, M. G., & Paivio, A. Pictures and words as stimulus and response items in paired-associate learning of young children. Journal of Experimental Child Psychology, 1968, 6, 231-240.

Dunn, L. M. Expanded manual: Peabody Picture Vocabulary Test. Minneapolis: American Guidance Service, Inc., 1965.

Durning, K. P. Preliminary assessment of the Navy Memory for Numbers Test. Unpublished master's thesis, San Diego State College, 1968.

Eells, K., et al. Intelligence and cultural differences. Chicago: University of Chicago Press, 1951.

Ehri, L. C., & Rohwer, W. D., Jr. Verbal facilitation of paired-associate learning as a function of syntactic and semantic relations. Paper presented at the annual meeting of the American Educational Research Association, 1969, Los Angeles.

Epstein, W., Rock, I., & Zuckerman, C. B. Meaning and familiarity in associative learning. Psychological Monographs, 1960, 74, No. 491.

Flanagan, J. C., & Cooley, W. W. Project TALENT one-year follow-up studies. Cooperative Research Project Number 2333. School of Education, University of Pittsburgh, 1966.

Frankel, F., & Cole, M. Measures of category clustering in free recall. Psychological Bulletin, in press.

Glasman, L. D. A social-class comparison of conceptual processes in children's free recall. Unpublished doctoral dissertation, University of California, 1968.

Green, R. B. SES differences on learning and ability tests in black children. Unpublished master's thesis, University of California, Berkeley, 1969.

Guilford, J. P. Fundamental Statistics in Psychology and Education. Third edition. New York: McGraw-Hill, 1956.

Guinagh, B. J. An experimental study of basic learning ability and intelligence in low socioeconomic populations. Unpublished doctoral dissertation, Michigan State University, 1969.

Horn, J. L. Organization of data on life-span development of human abilities. In Baltes, P. B., & Goulet, L. R. (Eds.), Life-span developmental psychology. New York: Academic Press, 1970.

Humphreys, L. G., & Dachler, H. P. Jensen's theory of intelligence. Journal of Educational Psychology, 1969, 60, 419-426. (a)

Humphreys, L. G., & Dachler, P. Jensen's theory of intelligence: A rebuttal. Journal of Educational Psychology, 1969, 60, 432-433. (b)

Jensen, A. R. Learning abilities in Mexican-American and Anglo-American children. California Journal of Educational Research, 1961, 12, 147-159.

Jensen, A. R. Learning abilities in retarded, average, and gifted children. Merrill-Palmer Quarterly, 1963, 9, 123-140.

Jensen, A. R. Individual differences in learning: Interference factor. Cooperative Research Project No. 1867, U. S. Office of Education, 1965.

Jensen, A. R. Social class, race, and genetics: Implications for education. American Educational Research Journal, 1968, 5, 1-42. (a)

Jensen, A. R. Patterns of mental ability and socioeconomic status. Proceedings of the National Academy of Sciences, 1968, 60, 1330-1337. (b)

Jensen, A. R. Another look at culture-fair testing. In Western Regional Conference on Testing Problems, Proceedings for 1968, "Measurement for Educational Planning." Berkeley, California: Educational Testing Service, Western Office, 1968, 50-104.

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123. (a)

Jensen, A. R. Jensen's theory of intelligence: A reply. Journal of Educational Psychology, 1969, 60, 427-431. (b)

Jensen, A. R., & Rohwer, W. D., Jr. Verbal mediation in paired-associate and serial learning. Journal of Verbal Learning and Verbal Behavior, 1963, 1, 346-352.

Jensen, A. R., & Rohwer, W. D., Jr. Syntactical mediation of serial and paired-associate learning as a function of age. Child Development, 1965, 36, 601-608.

Jensen, A. R., & Rohwer, W. D., Jr. Mental retardation, mental age, and learning rate. Journal of Educational Psychology, 1968, 59, 402-403.

Kushlick, A. Assessing the size of the problem of subnormality. In Meade, J. E., & Parkes, A. S. (Eds.), Genetic and environmental factors in human ability. New York: Plenum Press, 1966, 121-147.

Levin, J. R., & Rohwer, W. D., Jr. Verbal organization and the facilitation of serial learning. Journal of Educational Psychology, 1968, 59, 186-190.

Luria, A. R. The mind of a mnemonist: A little book about a vast memory. Translated from the Russian by Lynn Solotaroff. New York: Basic Books, 1968.

Marascuilo, L. A., & Levin, J. R. Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. American Educational Research Journal, 1970, 7, 397-421.

Martin, C. J. Associative learning strategies employed by deaf, blind, retarded and normal children. Educational Research Series, 1967, 38, 1-158.

Martin, C. J., Boersma, F. J., & Cox, D. L. A classification of associative strategies in paired-associate learning. Psychonomic Science, 1965, 3, 455-456.

Martin, C. J., Cox, D. L., & Boersma, F. J. The role of associative strategies in the acquisition of paired-associate material: An alternative approach to meaningfulness. Psychonomic Science, 1965, 3, 463-464.

Maxwell, J. Social implications of the 1947 Scottish Mental Survey. London: University of London Press, 1953.

Milgram, N. A. Retention of mediation set in paired-associate learning of normal children and retardates. Journal of Experimental Child Psychology, 1967, 5, 341-349. (a)

Milgram, N. A. Verbal context versus visual compound in paired-associate learning by children. Journal of Experimental Child Psychology, 1967, 5, 597-603. (b)

Montague, W. D., & Wearing, A. J. The complexity of natural language mediators and its relation to paired-associate learning. Psychonomic Science, 1967, 7, 135-136.

Morrison, D. F. Multivariate statistical methods. Chapter 5. New York: McGraw-Hill, 1967.

Nichols, R. C. Implications of racial differences in intelligence for educational research and practice. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1969.

Paivio, A. Meaning, mediation, and memory. Research Bulletin No. 48, Department of Psychology, The University of Western Ontario, London, Canada, 1967, 1-37.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A. On the functional significance of imagery. Psychological Bulletin, 1970, 73, 385-392.

Paivio, A., & Yuille, J. C. Mediation instructions and word attributes in paired-associate learning. Psychonomic Science, 1967, 8, 65-66.

Paivio, A., Yuille, J. C., & Smythe, P. C. Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. Canadian Journal of Psychology, 1966, 20, 362-377.

Palermo, D. S. Imagery in children's learning: Discussion. Psychological Bulletin, 1970, 73, 415-421.

Rapier, J. L. The learning abilities of normal and retarded children as a function of social class. Journal of Educational Psychology, 1968, 59, 102-110.

Raven, J. C. Guide to using the Coloured Progressive Matrices: sets A, Ab, B. London: H. K. Lewis & Co., Ltd., 1960.

Reese, H. W. Imagery in paired-associate learning in children. Journal of Experimental Child Psychology, 1965, 2, 290-296.

Reese, H. W. Imagery and contextual meaning. Psychological Bulletin, 1970, 73, 404-414.

Rohwer, W. D., Jr. Constraint, syntax, and meaning in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 541-547.

Rohwer, W. D., Jr. Social class differences in the role of linguistic structures in paired-associate learning: Elaboration and learning proficiency. Final report on U. S. Office of Education Basic Research Project No. 5-0605, Contract No. OE 6-10-273. November 1967.

Rohwer, W. D., Jr. Mental mnemonics in early learning. Teachers College Record, 1968, 70, 213-226.

Rohwer, W. D., Jr. Intelligence quotient versus learning quotient: Implications for elementary curricula. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1969.

Rohwer, W. D., Jr. Images and pictures in children's learning. Psychological Bulletin, 1970, 73, 393-403.

Rohwer, W. D., Jr. Mental elaboration and proficient learning. Minnesota Symposia on Child Psychology, Vol. 4, in press.

Rohwer, W. D., Jr., & Ammon, M. S. Locus of facilitation in noun-pair learning. Paper presented at the annual meeting of the Western Psychological Association, 1968, San Diego, California.

Rohwer, W. D., Jr., Ammon, M. S., Suzuki, N., & Levin, J. R. Populations differences in learning proficiency. Journal of Educational Psychology, in press.

Rohwer, W. D., Jr., & Levin, J. R. Action, meaning and stimulus selection in paired-associate learning. Journal of Verbal Learning and Verbal Behavior. 1968, 7, 137-141.

Rohwer, W. D., Jr., & Lynch, S. Semantic constraint in paired-associate learning. Journal of Educational Psychology, 1966, 57, 271-278.

Rohwer, W. D., Jr., & Lynch, S. Form class and intralist similarity in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 551-554.

Rohwer, W. D., Jr., & Lynch, S. Retardation, school strata and learning proficiency. American Journal of Mental Deficiency. 1968, 73, 91-96.

Rohwer, W. D., Jr., Lynch, S., Levin, J. R., & Suzuki, N. Pictorial and verbal factors in the efficient learning of paired associates. Journal of Educational Psychology, 1967, 58, 278-284.

Rohwer, W. D., Jr., Lynch, S., Levin, J. R., & Suzuki, N. Grade level, school strata, and learning efficiency. Journal of Educational Psychology, 1968, 59, 26-31.

Rohwer, W. D., Jr., Lynch, S., Suzuki, N., & Levin, J. R. Verbal and pictorial facilitation of paired-associate learning. Journal of Experimental Child Psychology, 1967, 5, 294-302.

Rohwer, W. D., Jr., Shuell, T. J., & Levin, J. R. Context effects in the initial storage and retrieval of noun pairs. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 796-801.

Runquist, W. N., & Farley, F. H. The use of mediators in the learning of verbal paired associates. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 280-285.

Semler, I. J., & Iscoe, I. Comparative and developmental study of the learning abilities of Negro and white children under four conditions. Journal of Educational Psychology, 1963, 54, 38-54.

Stevenson, H. W., Hale, G. A., Klein, R. E., & Miller, L. K. Interrelations and correlates in children's learning and problem solving. Monograph of the Society for Research in Child Development, 1968, 33 (No. 7).

Stodolsky, S. S., & Lesser, G. Learning patterns in the disadvantaged. Harvard Educational Review, 1967, 37, 546-593.

Suzuki, N. Noun-pair learning in children and adults: Deep structure and retrieval time. Unpublished doctoral dissertation, University of California, Berkeley, 1969.

Suzuki, N., & Rohwer, W. D., Jr. Verbal facilitation of paired-associate learning: Type of grammatical unit vs. connective form class. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 584-588.

Suzuki, N., & Rohwer, W. D., Jr. Deep structure in the noun-pair learning of children and adults, Child Development, 1969, 40, 912-919.

Talland, G. A. Deranged memory: A psychonomic study of the amnestic syndrome. New York: Academic Press, 1965.

Terman, L. M., & Merrill, Maud A. Stanford-Binet Intelligence Scale: Manual for the Third Revision, Form L-M. Boston: Houghton Mifflin, 1960.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers College Press, 1944.

Wallis, W. A., & Roberts, H. V. Statistics: A New Approach. New York: Free Press, 1957.

White, S. H. Evidence for a hierarchical arrangement of learning processes. In L. P. Lipsitt and C. C. Spiker (Eds.), Advances in child development and behavior. Vol. 2. New York: Academic Press, 1965.

Wilson, A. B. Social stratification and academic achievement. In A. H. Passow (Ed.), Education in depressed areas. New York: Teachers College, 1963.

Zeaman, D., & House, Betty J. The relation of IQ and learning. In Gagné, R. M. (Ed.), Learning and individual differences. Columbus, Ohio: Merrill, 1967.

Zigler, E. Familial mental retardation: A continuing silemma. Science, 1967, 155, 292-298.

Zigler, E. Developmental versus difference theories of mental retardation and the problem of motivation. American Journal of Mental Deficiency, 1969, 73, 536-556.

## Appendix A

## MEMORY FOR NUMBERS TEST

### Instructions to Subjects

This is a test to see how well you can remember numbers.

I will say some numbers and when I am through I want you to write the numbers in the boxes in the booklet in front of you.

Always listen carefully and try to get the numbers just right.

### Part I. Immediate Recall -- Practice Test

Now turn to page 2. I am going to say some numbers and after I am through saying them you will hear a bong. It sounds like this (bong). After you hear the bong -- but NOT before -- start writing the numbers I have just said in the boxes. Stop when you hear the bong again. Now remember, start writing when you hear the bong after I have said the numbers and stop writing when you hear the bong for the second time. After you stop writing hold your pencils up.

The first numbers I am going to say go into the boxes marked (a), the next numbers go into the boxes marked (b), and the last numbers go into the boxes marked (c). Now look at the boxes marked (a); listen carefully and write down as many of the numbers as you can remember when you hear the bong.

| | | | | |
|---|---|---|---|---|
| 725 | (bong-- | 13 secs.-- | bong) | |
| 831 | " | " | " | |
| 216 | " | " | | (End of practice) |

### Immediate Recall Test

Now turn to page 3. We will do the same thing as before except that there will be more numbers this time. I am going to say the numbers and after I am through saying them you will hear the bong. Start writing the numbers I have just said when you hear the bong, but not before. Stop writing when you hear the bong and hold your pencils up. I will then say more numbers and again you will write them down when you hear the bong. Listen carefully and get as many right as you can.

| | | | | |
|---|---|---|---|---|
| 3246 | (bong-- | 13 secs.-- | bong) | |
| 71493 | " | " | " | |
| 368279 | " | " | " | |
| 6817359 | " | " | " | |
| 97583264 | " | " | " | |
| 169248735 | " | " | | (End of practice) |

Now turn to page 4. We are going to do the same thing on this page as on the last page. Remember not to write before you hear the bong and to stop when you hear the bong again. Listen carefully and try to get as many numbers right as you can.

| | | | |
|---|---|---|---|
| 5472 | (bong--13 | secs.--bong) | |
| 63847 | " | " | " |
| 759326 | " | " | " |
| 3749521 | " | " | " |
| 43581926 | " | " | " |
| 368497152 | " | " | (End of page 4) |

Now turn to page 5. We will do the same thing as before. Try to get as many numbers right as you can when you hear the bong.

| | | | |
|---|---|---|---|
| 2794 | (bong--13 | secs.--bong) | |
| 97426 | " | " | " |
| 421637 | " | " | " |
| 7328649 | " | " | " |
| 38625714 | " | " | " |
| 271853649 | " | " | (End of Part I) |

## Part II. Repeated Series - Practice Test

Now we will do the second part of the test. This time, I will say the numbers three times. I will do it like this: first you will hear the numbers, then there will be a little noise. Then you will hear the same numbers again, then the little noise and the numbers for the third time. After I have said the numbers for the third time you will hear the bong. Start writing the numbers into the boxes when you hear the bong, but not before. Again, let us practice first. Turn to page 6. Remember, I am going to say the numbers three times. When you hear the bong after I have said the numbers three times, start writing the numbers you remember in the boxes and stop when you hear the bong again. Then hold up your pencils. Now listen carefully and get as many numbers right as you can.

| | | | | |
|---|---|---|---|---|
| 214 | (3 times) | (bong--13 | secs.--bong) | |
| 751 | " | " | " | " |
| 635 | " | " | " | (End of practice) |

## Repeated Series Test

Now turn to page 7. We will do the same thing we just practiced except that there will be more numbers this time. Remember, I am going to say the numbers three times, and when I am through saying them three times, you will hear the bong. Start writing the numbers in the boxes when you hear the bong but NOT before. Stop when you hear the bong again. Listen carefully and write as many numbers in the boxes as you can.

| | | | | |
|---|---|---|---|---|
| 6537 | (3 times) | (bong--13 | secs.--bong) | |
| 82196 | " | " | " | " |
| 769841 | " | " | " | " |
| 51638274 | " | " | " | " |
| 961354827 | " | " | " | (End of practice) |

Now turn to page 8. We will do the same thing on this page as on the last one. Try to get as many numbers in the boxes as you can.

| | | | | |
|---|---|---|---|---|
| 3269 | (3 times) | (bong--13 | secs.--bong) | |
| 42731 | " | " | " | " |
| 487214 | " | " | " | " |
| 3516829 | " | " | " | " |
| 87462139 | " | " | " | " |
| 752149638 | " | " | " | (End of page 8) |

Turn to page 9. We will do this page exactly like the one before. Listen carefully and write as many of the numbers down as you can.

| | | | | |
|---|---|---|---|---|
| 1482 | (3 times) | (bong--13 | secs.--bong) | |
| 29763 | " | " | " | " |
| 879362 | " | " | " | " |
| 6917243 | " | " | " | " |
| 32157948 | " | " | " | " |
| 294365817 | " | " | " | (End of Part II) |

## Part III. Delayed Recall -- Practice Test

Now we will do the last part of the test. This time, I will say some numbers and then there will be a short while when you hear nothing. Then you will hear the bong. I want you to hold up your pencils while I am saying the numbers and during the short while when you hear nothing. When you hear the bong, start writing the numbers you can remember in the boxes and stop writing when you hear the bong again. Then hold up your pencils until you hear the bong again after I have said more numbers. Let us practice first. Turn to page 10. Remember, I am going to say some numbers and then there will be a short while when you hear nothing. Hold up your pencils until you hear the bong and write the numbers I have just said until you hear the bong again. Listen carefully and write down as many numbers as you can remember.

| | | | | |
|---|---|---|---|---|
| 639 + 10 secs. | (bong--13 | secs.--bong) | | |
| 241 + " | " | " | " | |
| 824 + " | " | " | (End of practice) | |

## Delayed Recall Test

Now turn to page 11. We will do the same thing that we just practiced except that there will be more nuubers this time. Remember, I am going to say some numbers and then there will be a short while when you hear nothing. Hold your pencils up until you hear the bong and then write down all the numbers you remember. Stop when you hear the bong again. Listen carefully.

| | | | | | |
|---|---|---|---|---|---|
| 9165 | + 10 secs, | (bong--13 | secs.--bong) | | |
| 36471 | " | " | " | " | |
| 761358 | " | " | " | " | |
| 9183574 | " | " | " | " | |
| 15279486 | " | " | " | " | |
| 375942186 | " | " | " | (End of page 11) | |

Now turn to page 12. We will do the same thing as we did on the last page. Try to get as many of the numbers right as you can. Listen carefully.

| | | | | | |
|---|---|---|---|---|---|
| 7495 | + 10 secs. | (bong--13 | secs.--bong) | | |
| 47216 | " | " | " | " | |
| 586974 | " | " | " | " | |
| 1963742 | " | " | " | " | |
| 86153794 | " | " | " | " | |
| 271854963 | " | " | " | (End of page 12) | |

And now turn to page 13. Again, we will do the same thing as on the last page. Listen carefully and write down as many numbers as you can remember.

| | | | | | |
|---|---|---|---|---|---|
| 2985 | + 10 secs. | (bong--13 | secs.--bong) | | |
| 95738 | " | " | " | " | |
| 376842 | " | " | " | " | |
| 4381629 | " | " | " | " | |
| 29758436 | " | " | " | " | |
| 386597241 | " | " | " | (End of Part III) | |

End of Test

Appendix B

# MEMORY FOR NUMBERS TEST

NAME ______________________________
First Last

SCHOOL __________________ Grade ________

TEACHER ____________________________

---

1 ____

2 ____

3 ____

---

4 ____

5 ____

6 ____

---

7 ____

8 ____

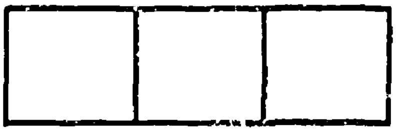
9 ____

| | I | R | D |
|---|---|---|---|
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |

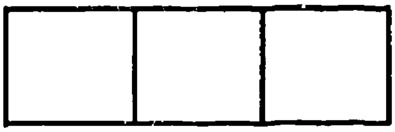
Appendix B

# PRACTICE SERIES

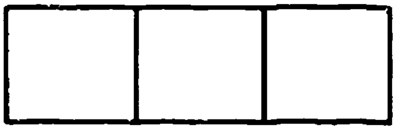
a



b



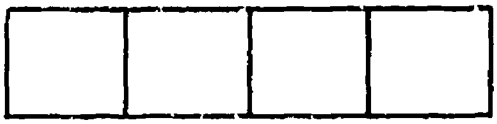
c



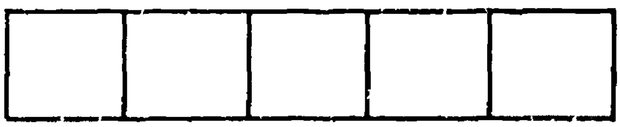
Appendix B

# TEST SERIES

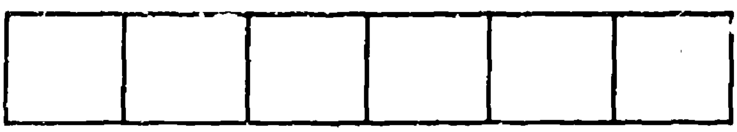
a



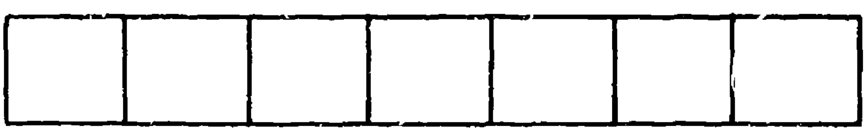
b



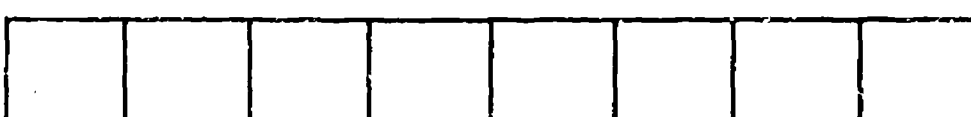
c



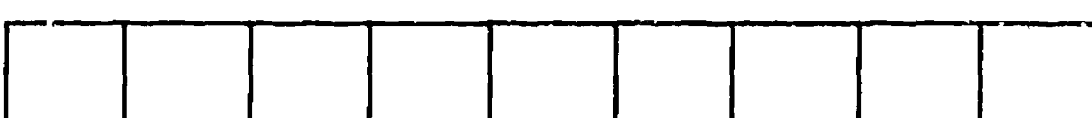
d



e



f



# Listening-Attention Test

Name ______________________________
Last First

School ____________________ Grade ______

Teacher ______________________________

---

## Practice Series

| a | | b | |
|---|---|---|---|
| 2 | 3 | 7 | 5 |
| 1 | 4 | 1 | 6 |
| 3 | 8 | 6 | 7 |
| 9 | 7 | 2 | 6 |
| 4 | 3 | 1 | 2 |

1 ________
2 ________
3 ________
4 ________
5 ________
6 ________
7 ________
8 ________
9 ________
10 ________

| A | B | C | D | E |
|---|---|---|---|---|
| 2 6 | 6 8 | 3 7 | 5 9 | 2 5 |
| 1 2 | 2 5 | 6 5 | 4 3 | 1 8 |
| 6 2 | 4 8 | 4 2 | 5 1 | 7 4 |
| 8 7 | 1 7 | 9 5 | 1 8 | 9 2 |
| 2 5 | 4 9 | 1 2 | 5 9 | 1 9 |
| 8 2 | 9 3 | 5 4 | 7 8 | 3 1 |
| 3 4 | 3 1 | 7 1 | 8 1 | 9 4 |
| 5 3 | 2 6 | 8 2 | 9 2 | 3 8 |
| 2 5 | 2 3 | 1 7 | 6 4 | 2 4 |
| 9 8 | 6 3 | 4 9 | 1 8 | 2 7 |

| F | | G | | H | | I | | J | |
|---|---|---|---|---|---|---|---|---|---|
| 7 | 2 | 8 | 1 | 9 | 4 | 3 | 6 | 8 | 9 |
| 4 | 2 | 4 | 6 | 7 | 1 | 3 | 2 | 5 | 2 |
| 8 | 3 | 2 | 9 | 2 | 8 | 6 | 2 | 3 | 5 |
| 4 | 9 | 1 | 8 | 1 | 7 | 1 | 3 | 4 | 3 |
| 1 | 3 | 6 | 7 | 4 | 5 | 3 | 9 | 2 | 8 |
| 9 | 1 | 9 | 5 | 2 | 1 | 9 | 4 | 5 | 2 |
| 2 | 9 | 8 | 1 | 5 | 9 | 7 | 1 | 7 | 8 |
| 4 | 7 | 1 | 5 | 2 | 4 | 8 | 4 | 2 | 6 |
| 8 | 1 | 3 | 4 | 5 | 6 | 5 | 2 | 2 | 1 |
| 5 | 2 | 9 | 5 | 7 | 3 | 8 | 6 | 6 | 2 |

Appendix D

## MAKING X's TEST

### Instructions

1. Have children fill out names, etc.

2. Tester should put a series of boxes on the blackboard in full view of class.

3. Say: "I want to see how well you can make X's in boxes, like this." (Demonstrate on blackboard.)

4. Then say: "Now, let's practice it in these boxes on the white page." (Point to place on page on your copy.)

5. Look around to see that everyone made X's in the practice boxes.

6. Say: "Now turn to the pink page. Don't write until I tell you to. Hold your pencils UP. When I say "Go", make X's in the boxes. Do your best, and stop when I say Stop! All right, get ready, GO!"

   Start stopwatch with word GO!

   AFTER exactly 1 minute, 30 seconds (total of 90 seconds) say: "STOP! PENCILS UP!

7. And then say: "Now turn to the YELLOW page. Keep Pencils up. This time I want you to work as fast as you can. Try your best to make more X's than you did the first time. Get ready, set, GO!" (Start stopwatch with word GO!)

   AFTER exactly 1 minute, 30 seconds, say: "STOP! PENCILS UP!"

   Collect test booklets immediately.

# Making Xs Test

Name ______________________________
Last Name First Name

School ________________________ Grade __________

Teacher ____________________________

## Practice Series

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |

1. ________
2. ________
3. ________
4. ________
5. ________
6. ________
7. ________
8. ________

# Appendix D